WARREN QUINN

The Right to Threaten and the Right to Punish

Most of us feel certain that punishment is, in many cases, fully justified. But as to the nature of the justification we are perplexed and uncertain. I do not refer here to punishment within the family. Parents are natural educators morally charged with the task of turning their young dependents into civilized adults, and they need, common sense insists, the possibility of punishing to succeed. But civic punishment, in which one adult is made to suffer for his past wrongdoing by other adults who officially represent the community, raises different problems.[1] It is not, of course, that we doubt its utility. Common sense urges, no less here than in the case of family discipline, that some form of civic punishment is necessary for a decent social order. The difficulty lies rather in the question of authority or right. For on the modern liberal view, adult criminals are not dependents of the community, and the community is not assigned the moral task of forming or improving their characters. How then does its right to punish them arise?

The major source of theoretical difficulty here is the fact that the restrictions, confinements, and deprivations of property and life that make up standard civic punishments would, if imposed in nonpenal contexts, be opposed by various important moral rights of liberty, life, and property. If these evils did not come by way of just punishment, a person subjected

I would like to thank Philippa Foot, Miles Morgan, Stephen Munzer, David Sachs, the Editors of *Philosophy & Public Affairs*, and, especially, Rogers Albritton for helpful comments on earlier drafts.

1. I would simply call this kind of punishment "legal" if I were certain that having a code of punishable behavior and designated authorities to punish always, even in very simple societies, adds up to having a legal system. In any case, I shall help myself to the word "crime" for the kind of thing that is properly punished by this type of punishment.

to them could object to his treatment in ways that would have serious moral weight.[2] To understand how civic punishment can be morally justified, therefore, we must first understand why these familiar rights do not stand in its way. There is, it is important to note, a parallel theoretical issue in the case of self-defense. In defending oneself against an unjust attack, one may put the attacker at some risk of being harmed or even killed, a risk that one could not create in most other contexts without violating some of his rights. Here, however, we feel that we can explain why an attacker's objections to such a defense do not count. How could morality first declare that certain aggressions would be serious violations of our rights and then extend to these aggressions an immunity from interference comparable to that which it assigns to innocent actions? But no parallel explanation seems available in the case of punishment. We cannot punish a crime until it is beyond influence.

If criminals fully retained their ordinary rights to liberty, life, and property, these rights would either raise a morally decisive barrier against punishment, in which case it would *violate* them, or they would create an obstacle that a case for punishment could override, in which case it could *justifiably infringe* them. Justified punishment could not, of course, violate rights. But perhaps it could be thought to infringe them. This idea invites us to include proper punishment in the class of actions, such as the expropriation of private property in time of national emergency, in which we regretfully but justifiably encroach upon someone's rights in order to prevent some evil. But, upon reflection, this assimilation appears doubtful.[3] For justified infringement of rights is a special moral circumstance creating special moral demands not present in the case of punishment. When one has harmed someone in the course of justifiably infringing his moral rights, it is always appropriate and sometimes required that one express regret and offer compensation. But when punishment is fully justified, expressing regret seems, at most, morally optional, and making compensation seems definitely out of place. We do not feel that a properly punished criminal is entitled to either.

2. Having a moral right, in the sense I intend throughout this article, is having a moral status in virtue of which one's objecting (or objecting that could be done on one's behalf) to what others might (or might not) do creates at least a prima facie obligation that they refrain from doing (or not doing) it.

3. Herbert Morris expresses such a doubt in "The Status of Rights," *Ethics* 92 (Oct. 1983): 45, and also offers some plausible objections to received conceptions of infringement.

It thus appears that a morally justifiable practice of punishment can neither violate *nor* infringe a criminal's rights. And therefore the central problem for any moral theory that takes both punishment and rights seriously is to show how this can be so despite the fact that in punishing we subject people to treatment that in other contexts would violate or at least infringe their rights. This is not, of course, the only interesting moral question that can be raised about punishment. A particular punishment might be unwise or unkind without violating any rights. But, as a natural working hypothesis, I shall assume that when punishment is *unjustified* because of the evil imposed on the punished person, it is unjustified as a violation of one or another of his rights. In discussing the question of rights, therefore, I shall often speak as if I am discussing the general question of justification. If this assumption should prove false, my argument will bear only on the specific question of the right to punish, that is, the question how punishment can be shown not to violate a punished person's moral rights.

The way I shall set about answering this question differs from the way state-of-nature theorists often proceed. They typically begin by assuming a nonproblematic right of private punishment (or, as I shall say, retaliation) in a state of nature. And from this they infer that the central philosophical problem for the theory of civic punishment is to show how such a right can be preempted by the state.[4] Now, accepting their idea that the right to civic punishment must somehow arise out of a more fundamental right to private retaliation, I agree that this is an important problem. But even in the case of retaliation, there is a more basic question. To give an adequate account of justified retaliation in a state of nature we must be able to explain why an offender's moral rights do not stand in opposition to the evil inflicted upon him. Justified retaliation raises, therefore, the same fundamental question raised by civic punishment. And since it is this question that I wish to answer, I shall avoid altogether the less basic question how a community can rightfully forbid private retaliation and force its members to accept instead the protections of civic punishment. I shall do this by restricting the discussion to communities whose members prefer the protections afforded by their practices of civic punishment to those they could hope to gain by threatening retaliation.[5]

4. The problem that occupies Robert Nozick in Part I of *Anarchy, State and Utopia* (New York: Basic Books, 1974).

5. To further lighten my burden, I shall also limit the discussion to punishment for acts

Before presenting my own account, it may be well to consider briefly how some familiar theories would address the problem of the right to punish. If any of these theories had a fully satisfying account of this right, there would be no need to continue looking. We may begin with consequentialist theories. Act consequentialism justifies an individual act of punishment by direct reference to its results. Among the most useful of these results is strengthening the deterrent effect on others of the ongoing threat. Since this consequence is a conspicuous social benefit, a plausible consequentialism cannot set it aside as irrelevant to the question of justification. Indeed, it is by reference to this kind of benefit that, under this theory, most justified acts of punishment receive the major part of their justification.

In rule consequentialism this kind of benefit enters into the justification of the practice of punishment as a whole and therefore also into the justification of acts of punishment. For a practice, insofar as it can be consequentially justified, includes those events that help constitute its existence, and these constituents must, in the present case, include the particular acts of punishment that would not otherwise occur. These acts are therefore a large part of what is justified when the practice as a whole is justified. And their deterrent effects on others are a large part of what does the justifying. It thus appears that both kinds of consequentialists are ready to justify punishments, at least in part, by reference to their deterrent effects on others. And even in rule consequentialism, the deterrent effects on others of a *particular* act of punishment may, in theory, tip the balance so as to justify the practice that contains it and, therefore, serve to justify the act itself.

To apply consequentialist theory of justification to the question of rights, we seem driven to the following result: Punished persons have no rights that stand against their punishment because, in part, punishing them is so often useful in helping to deter others from committing crimes. But I, for one, find this answer deeply disturbing. There may indeed be situations in which utility decides the presence or absence of moral rights. But to justify punishment in this way is to say that properly punished people lack the relevant rights because, in large part, they make such useful object lessons for others. In no other case, however, do we suppose

---

that clearly violate people's public moral rights, acts such as murder, theft, assault, and fraud.

that ordinary rights to liberty and life fail to apply *because* their application would stand in the way of some socially profitable use of people.[6] Our rights, by their very nature, are kinds of moral properties that resist such attempts to justify incursions upon them. The most that can follow from an appeal to the general utility of using people is that any rights that stand in its way may be justifiably infringed. But it is implausible, as we have seen, to regard punishment as justified infringement.

A deterrent theorist might escape this kind of objection by restricting his appeal to the deterrent effects of punishment on the person who is punished. Punishing in an attempt to make a criminal's future behavior morally acceptable could not naturally be construed as making use of him. And in aiming at the social utility that would result from this improvement rather than at the punished person's edification, punishers would not be liable to the charge of paternalism. The trouble is that such an account restricts the class of cases in which punishment can be justified and thus unreasonably restricts the kinds of liabilities to punishment that can be created. Suppose, for example, that we discover that a certain type of person is psychologically capable of committing only one murder. Such a person need never murder at all, but if he does murder once he will never murder again. If such people could be identified, we could not on this view rightly punish them for murder since we could not justify this punishment as a way of keeping them from committing future murders. And this seems to imply that we could not, except as a bluff, have the threat of punishment for murder stand against them in our penal code. But this implication seems absurd. Surely we might rightly make them liable to punishment in hope of deterring the single murder that each is capable of committing.[7]

We must also consider the family of retributivist theories of punishment. Often, caught up in establishing a supposed duty to punish, retributivists do not directly address the question of right. But when they

6. The military draft in wartime may seem an exception to this claim. But even here it would be odd to argue directly from social need to the conclusion that people have no right not to be forced into military service. Claims about what is *owed* the community in return for alleged benefits received from it are at least implicit in all common-sense justifications of conscription. Moreover, it may be possible to view conscription as a justifiable infringement of liberty rights and to interpret the various forms of preferential treatment accorded to veterans as forms of compensation.

7. This objection, suitably modified, can also be used against other educative and reformist conceptions of punishment.

do, some retributivists invoke the idea of *forfeiture*. The rights that would otherwise have barred us from doing the sorts of thing we do in punishing, for instance, depriving the criminal of liberty or life, have been forfeited by his own behavior. These rights are seen as conditional and, therefore, liable to deteriorate or disappear unless preserved by a certain moral prudence. This conditionality can be seen as a basic feature of the operation of natural moral law that provides an independently intelligible "clearing of the way" for retribution.

The appeal to forfeiture as an independently intelligible moral mechanism is, however, problematic. The proper authorities are entitled to punish Jones, a generally decent young man who has foolishly stolen Smith's car, by depriving him of up to the amount of liberty forfeited in the theft. But suppose that before any such punishment takes place, Smith, for reasons having nothing whatever to do with the theft, kidnaps Jones and deprives him of exactly that amount of liberty. In this situation it is natural to suppose that Smith not only wrongs Jones but specifically violates his right to liberty. Perhaps this is because Jones forfeits his right to the community as a whole and not to Smith in particular. But suppose that the community in which Jones lives has the unjust practice of seizing and confining political dissenters. And suppose that shortly after his crime Jones, who also happens to be a dissenter, is officially seized and, for a time, quarantined to prevent the spread of his political views (views having nothing to do with his theft). Again, we would naturally suppose that Jones's right to liberty had been violated by his community, even if he were confined only for a period that would constitute an acceptable punishment for his theft and were never punished thereafter. But surely all this strongly suggests that the conditionality of Jones's right to liberty (the conditionality invoked by the doctrine of forfeiture) makes essential reference to punishment.[8] Jones has not forfeited his right without qual-

8. To deny this would be to adopt a theory of forfeiture according to which Jones's crime has to some extent made him an *outlaw*, someone whose basic moral rights do not stand in the way of a certain amount of ill-treatment whether or not it comes by way of punishment. Someone who held this idea of forefeiture might try to show that it could be restricted in various ways by moral and legal conventions. He might in this way hope to account for the fact that, in our own moral and legal systems, Jones forfeits his conventionally specified liberty right only in that he may be punished. On such a view, forefeiture would be restricted only in some systems of social morality; in other morally acceptable systems outlawry in some degree would be the regular consequence of crime. It is this last suggestion that I find disturbing.

ification, he has forfeited it in that he may be subjected to a certain penalty (presumably the proper penalty for the crime) by certain people (presumably those with the right to punish him). It seems, therefore, that the idea of forfeiture in this kind of case comes to no more than the idea that the criminal's rights do not in fact stand in the way of his being punished. The appeal to forfeiture does not, as it first seemed, provide an explanation of why this is so.[9]

It is sometimes thought that the force of an appeal to forfeiture lies in the moral necessity of reciprocating respect for rights. But it is not clear how this necessity can help explain why the loss of rights arising from nonreciprocation focuses so precisely on punishment. Moreover, such an account is hard pressed in other ways. If respect is treated as an *attitude*, then many people who never steal may have as little respect for property as Jones, who may have uncharacteristically succumbed to an unusually strong temptation.[10] Yet these others do not, in virtue of their attitude, forfeit any rights. If, on the other hand, respect is taken to refer only to actions, then it is not clear what role the idea of reciprocation is to be assigned. Suppose Smith wrongly takes Jones's car at the very same time

9. Forfeiture of the kind we are discussing might be distingushed from forfeiture whose very possibility is created by a contract that both creates a right and specifies the precise ways in which it may be lost. Some philosophers might appeal to forfeiture of the latter kind by construing the right to punish as deriving from a hypothetical social contract that both creates and limits various social rights. On such a picture, our natural disapproval of, for example, murder and assault, would lead us to design our rights in a way that stipulated forfeiture for such acts. And certain other natural desiderata would lead us to specify that this forfeiture be to the community for the specific purpose of punishment. While tempting, this kind of account raises some difficult questions. First, there is the familiar problem of the actual moral force of a hypothetical agreement. Second, it would have to be shown that such a view can make sense of the morally intuitive upper limits on punishment. And this would be difficult, I think, even if the hypothetical contractors were trying to minimize the likelihood of the worst things that can happen to them. For in the design of their future practice they will focus not so much on individual crimes, and how best to deal with them, as on the bearing of alternative possible practices on their lifetime prospects. And since one very bad, but empirically possible, lifetime prospect is to be a *repeated* victim of a certain kind of crime, it may be reasonable to design the practice so as to tolerate, under certain empirically possible situations, Draconian penalties. Finally, this picture of the right to punish seems to give no account of the right of retaliation in a state of nature—a right which, to my mind, not only exists but raises the same basic theoretical problem as does the right to punish.

10. There is also the difficulty of explaining why Jones doesn't lose his rights only for the period in which he remains disrespectful of others' rights. Without such an explanation it will not be clear with what right we punish people who have reformed in the interval between crime and punishment.

that Jones takes his. They apparently reciprocate the same degree of respect for each other's rights. But surely this does not mean that neither may be punished. Crimes may be punishable even though everyone commits about the same number of them.

A retributivist may, however, omit any appeal to forfeiture in his account of the right to punish. He may argue that the special moral character of retribution, its status under justice as something *deserved*, demands that morality make room for it.[11] Morality would, on such a view, be internally inconsistent if it fully extended ordinary rights into penal contexts. Since it is not inconsistent, a person who gets what he deserves cannot object by appealing to any moral right. When retributivism is thus conceived, evaluation of its account of the right to punish must focus on the moral credentials of retribution itself. Since a critical examination of these credentials would take me beyond my present purpose, I shall make only two brief observations. First, it seems clear that retributivism is burdened by a prima facie mystery. The idea that it is just (and, therefore, in some sense morally good) to harm someone's interests simply because he has wrongly harmed someone else's interests is, when considered in the cold light of reason, hard to understand. Second, one may doubt that an appeal to particular moral intuitions can help to dispel the mystery. For while we would often be inclined to assent to the claim that a particular person deserves to suffer for what he has done, the thought behind our inclination may not be adequately expressed in the precise words we accept. Our underlying thought may simply be that the criminal ought to be punished for his crime and that his punishment will be justified not by its effects but by the fact of the crime itself. But this intuition can

11. Some quasi-retributivist conceptions that bring punishment under the heading of rectification also do without forfeiture. Versions of these views can be found in Herbert Morris's "Persons and Punishment," *The Monist* 52, no. 4 (October 1968): 475–501, and in Jeffrie Murphy's *Retribution, Justice and Therapy* (Boston: Reidel, 1979), pp. 73–115 and 223–49. Such accounts, while attractive in many ways, face a number of problems. If punishment is modeled on the payment of a debt or the cancellation of an illicit liberty, it must be explained why the matter cannot be set right, voluntarily, in other ways. Moreover, in clear cases of rectification, there is the possibility of transferring the misappropriated property or power either to its rightful owner or to someone with a better claim to it than the wrongful possessor has. In punishment, however, there is a "taking away" from the criminal without any obvious transfer of what is taken away to anyone else. For some other criticisms of these views see Richard Burgh's "Do the Guilty Deserve Punishment?," *Journal of Philosophy* 79, no. 4 (April 1982): 193–210.

be valid in the nonretributivist account of punishment that I shall now present.

This conception, unlike those we have considered, gives equal attention to two temporally distinct components of the practice of punishment. The first is establishing the real risk of punishment, creating serious *threats* of punishment designed to deter crime. The second is, of course, the actual *punishing* of those who have ignored the threats. According to this conception, the standard theories err in assuming that the right to threaten punishment derives from the anticipation of an independently intelligible right to punish. The central idea of this conception is, in contrast, that the right to make people liable to punishment is the *ground* of our right to punish.[12] Another way to put this claim is to say that according to conventional theories one cannot object to being subject to the threat of punishment because one will not, if one commits the offense, be able to object to being punished, whereas on the present view, it is because one could not object to the threat in the first place that one cannot, later, object to being punished.

To create (or establish) a threat against x, in the quasi-technical sense that I intend in this discussion, is *first, deliberately to create a real risk that x will suffer a certain evil if he does or omits a certain specified action and second, to warn x of the existence of this risk, where by these means x may possibly be deterred from the act or omission.*[13] I mean this quasi-technical use to include the making of ordinary sincere threats.[14] In this most simple case the conditional danger lies in the intention of the threatener to carry out the threat. But in typical practices of punishment the conditional danger derives from an already existing "machinery" of the law. This "machinery" is, in large part, made up of the dispositions of various functionaries to make their assigned contributions to the proc-

12. Thomas Hurka advances this idea in "Rights and Capital Punishment," *Dialogue* 21, no. 4 (December 1982): 649. (But he seems to take it back or at least to modify it on p. 659, where he asserts that punishment, by which I take him to mean punishing, is impermissible unless *it* promotes the social good.) Hurka suggests a view of punishment in many respects like the one I am advancing, but he defends it on p. 650 with what seems to me a very dubious libertarian argument.

13. As I mean it, the warning condition is satisfied where a general risk that applies to x is publicized but x does not, through his own fault, become aware of it and in cases where x knows of the risk independently of any explicit warning that might be given.

14. But mere bluffs and threats in which the threatener hasn't really decided whether he would carry out the threat are excluded.

ess by which guilty persons come to be punished. But I shall also be considering a kind of general threat in which the danger is created by activating artificial devices. To speak of a threat in such a case will further extend the ordinary notion, which normally includes the idea that the threatened evil will come by way of some intentional human action.

When we create threats of punishment we are, according to the theory I wish to develop, justified by our rights of *self-protection*.[15] It is morally legitimate to create these threats because it is morally legitimate to try to protect ourselves in this way against violations of our moral rights. Viewed the other way round, we cannot object to certain deterrent threats of punishment that stand against us because others have a right to try to protect themselves from us by these means. The theory asserts that a practice of punishment is at its moral core a practice of self-protective threats.

I shall try to make this two-stage conception of punishment clearer and more plausible in the course of the discussion. But even from the present sketch one can see that it is in some ways like both standard deterrent and standard retributive theories. Like the former, it refers the justification of punishment to the goal of prevention. But unlike them, it does not try to justify *acts* of punishment as means to that end.[16] Only the prior threats are justified in this way. Like familiar retributive theories, it is backward-looking in its account of the right to punish, but unlike them it invokes no primitive notion of desert. Instead it explains the right to punish by reference to the right to establish the original threat.

This conception raises two different questions. First, whether the right to create the threat of punishment is, as I claim, grounded in a right of self-protection. Second, whether the right to punish is, as I claim, derivative from the right to establish the threat. The first question is not totally

15. In "The Doomsday Machine: Proportionality, Punishment and Prevention," *The Monist* 63, no. 2 (April 1980), Lawrence Alexander distinguishes two different practices within what we call punishment, one of which rests on what I call rights of self-protection.

16. In this respect it may seem to be like the well-known mixed views of punishment put forward by John Rawls in "Two Concepts of Rules," *Philosophical Review* 44 (1955): 3–32, and H.L.A. Hart in "Prolegomenon to the Principles of Punishment" in Hart, *Punishment and Responsibility* (New York: Oxford University Press, 1968). In their rule-consequentialist aspects, however, both these views (if my previous argument was correct) justify acts of punishment, taken collectively, by reference to their collective preventive effects. The crucial distinction in these theories is between whole (practice) and part (act) while in the present theory it is between the earlier threat and the later punishment.

independent of the second, and we therefore need a strategy for resolving all of its independently resolvable parts. The strategy I have chosen is to construct an imaginary practice of threatening in which all threatened evils are to be delivered by fully automatic devices. This twist allows us to examine the moral basis for making threats in abstraction from the question of the specific right to carry them out. The kind of automated practice I will construct is, of course, meant to be *purely self-protective* in its function and moral ground. I shall examine such a practice in Section I, where I shall not presuppose that it is, apart from its automatic character, the moral equivalent of punishment. I shall compare the two practices in Section II, where I shall argue that they are indistinguishable in the distribution of acceptable penalties. I will try to show that, apart from the question of the right to carry out actual punishment, there is nothing in the workings of a justified practice of punishment that prevents us from seeing it as a practice of self-protective threats. I shall confront the final difficulty in Section III, where I shall try to show how the right to punish can derive from the right to create the earlier threat.

I

Let us imagine ourselves existing at some time in the future when, our social structures having been destroyed by earlier upheavals, we come together to form a new community. Being scientifically very advanced, although no more moral or prudent than we were, we are together capable of making fantastically complex devices that can (at least as well as we can) detect wrongdoing in our new community, identify and apprehend those who are responsible, establish their guilt, and subject them to incarcerations (and perhaps other evils) that I shall call mechanical-punishments, or m-punishments for short.[17] Let us further imagine that we have lost whatever taste we once had for retribution and are interested only in protection. The devices attract us, therefore, in their deterrent rather than in their retributive capacities. Furthermore, we are not particularly concerned with the theoretical question whether using the devices for protection would constitute a new form of civic punishment.

17. The antecedents of these devices in the current literature are Lawrence Alexander's Doomsday Machines which differ from our devices in offering only one drastic penalty for any crime, and James Buchanan's automatic enforcing agents in *The Limits of Liberty* (Chicago: University of Chicago Press, 1975), p. 95.

We are prepared, however, to acknowledge that there may be limits within the morality of self-protection that resemble limits on justifiable punishment. And we are therefore glad of the sophistication of the devices, which may not only be programmed to mete out different m-punishments for different crimes, but may also be set, in virtue of their remarkable (but not infallible) ability to determine the state of mind in which someone committed a crime, to withhold m-punishments in kinds of cases that we do not wish to bring under the threats. We are also glad that the devices can be so conservatively programmed that they will never be "fooled" into m-punishing the innocent. For, as it happens, we prefer this form of safety even though it means that the devices may sometimes fail to identify the guilty, a possibility made even more likely by our further wish that they be programmed to "pay" scrupulous attention to civil liberties in their criminal "investigations."[18]

It is not surprising that we are drawn to the prospect of deterring crime by means of the general threats posed by these devices. We are attracted, to be sure, by their efficiency and powers of discrimination (which I am supposing to be at least as great as ours), but we are even more pleased to leave the unpleasant business of enforcement to nonhuman enforcers. For it must be remembered that these automatic devices, while marvelously sophisticated, are not persons responsible for authentic moral choices. Should we choose to use them, their operation would involve only that choice and the choices that determined their design and program. To insure the completeness of this desirable isolation from human control let us also suppose that once the devices have been activated for a certain predetermined period, they cannot then be interfered with. For the choice whether or not to stop or alter them could, at some point, be tantamount to the choice whether or not someone will be m-punished.

The fact that our only choices are initial ones seems to throw new light on the question of our right to protect ourselves by way of threats. Without the devices, an effective system of deterrent threats can exist only if some of us are prepared to make independent choices to carry them out. And these independent acts must, it naturally appears, be justified in one of

18. Throughout this article I shall simply ignore the difficulty of justifying a practice that, like all actual practices, sometimes punishes the innocent. And by imagining devices that are programmed to operate under all the principled constraints that would govern the best human punishers, I hope to avoid any objection to their automatic and artificial character.

the familiar but unsatisfying ways already considered.[19] With the devices, however, there are no such independent choices. There is instead an initial choice to establish an ongoing deterrent threat, where it is foreseen that this will, in all probability, cause m-punishments to occur. In this situation, our choice to bring about future m-punishments is *derivative*, an unavoidable consequence of our more basic choice to set up the threat. That this is so is shown by the fact that we would not choose to create the threat if we foresaw that m-punishments would occur without protecting us, but that we would choose to create the threat if we foresaw that it would be so effective that m-punishments would never occur. And it is morally significant that our choice to bring about m-punishments is, in this sense, derivative. For derivative choices may sometimes be justified by that which can be said for the basic choices to which they attach. In our new situation, therefore, we may be able to see how to justify bringing about m-punishments even if we cannot see how to justify independent choices to carry out threats that have failed.

This restructuring of the problem would not be available if we were retributivists using the devices as surrogate retributors. For then each m-punishment would be a kind of end for us. But our actual end is nothing other than protection from crime. And it is the threat of m-punishment rather than m-punishment itself that we call upon this end to justify. It is important, however, to see that our choice is to create a *real* and not a deceptive, threat. A deceptive threat may have the advantage of not leading to any m-punishments, but it will not protect us as well as a real threat. For some people will have to know of the deception. And, in the scientifically advanced community we are imagining, others will surely suspect it. More important, such deception would be morally insupportable. It is one thing for a private individual to protect himself by bluffing. But it is an altogether different thing for civic authorities, acting in their official capacities, to practice wholesale deception in a matter as vital as this to each citizen's interest. We therefore have both practical and moral reasons to create a real threat if we create any.

It does not, however, follow from the fact that our threat is real and will, therefore, almost certainly lead to m-punishments that our initial justification must appeal to the anticipated deterrent effects of those m-

19. I shall argue in Section III that this appearance is deceptive, and that the justificatory structure of an actual punishment system is really like that of a system of m-punishment.

punishments. If such an appeal were necessary, then the future m-punishments themselves, along with the threat of them, would enter our scheme of justification as means of protection. But the justifiction we seek makes no such appeal. We see that each m-punishment will occur as the result of the previous existence of a real threat, and we insist that each such prior threat be completely justified by reference to the protection that *it* can be expected to create.

To see how this justification works, we may begin by considering the initial period from the moment of activation up through the occurrence of the first crime that the devices will subsequently m-punish. We do not know exactly how long or short this period will be. But we have good empirical grounds for believing that, given human nature, it cannot be very long. The activators must therefore ask themselves whether they would be justified in establishing the threat (with its risk of giving rise to m-punishments) for any stretch of time that might realistically constitute this initial period even if the deterrent force of the threat were not to be reinforced by the publicized occurrence of any m-punishment. This is to insist on a justification for activating the devices for any such stretch that appeals only to protection that would result from the publicized fact of activation itself (from the general belief that the devices will work) and from possible artificial demonstrations of their effectiveness.

If the protection created by these factors alone would justify establishing the threat for any such duration, then the first m-punishment would clearly be justified, not as a means to later protection, but as an unavoidable empirical consequence of our having enjoyed an earlier protection.[20] And each subsequent m-punishment would presumably be justified in the same manner, by reference to the period of threatening that preceded it. Of course, there may come a time when the deterrent effects of publicized m-punishments become essential to the continued justifiability of the ongoing threat. If, fantastically, all m-punishments were kept secret, some would-be criminals might eventually cease to believe in the reality of m-punishment. But given that each actual nonsecret m-punishment is justified by reference to the threat that preceded it, each

20. Speaking of the justification of m-punishments is, of course, shorthand for speaking of the justification of the original activation insofar as it was foreseeable that they would result. As already noted, m-punishments are not real actions, and the devices are not real agents. So, strictly speaking, m-punishments are not the sorts of things that can be morally justified or unjustified.

may be allowed to contribute its deterrent effects to the case for the continuing threat. For it does so as something already justified quite independently of those effects.

The key to this scheme is the fact that activation for a shorter term never depends for its justification on activation for a longer term. The first, second, third, etc. m-punishments are therefore each justified as empirically unavoidable costs of and not as producers of protection. Each m-punishment is seen as the byproduct of a period to which it does not contribute any deterrence. And it is nothing other than the fact that there is no separate choice to bring about the m-punishments, that our choices are tied together in a single initial choice to activate and create the standing threat, that makes this pattern of justification stand out.

On this conception, everything depends on our initial right to protect ourselves by placing would-be criminals under real threats. This right is akin to, but in some ways different from, the right of self-defense and the right to construct protective barriers. It allows us to make offenses less tempting by attaching to them the real prospect of costs, in this case costs that may, as a matter of the operation of automatic devices, follow crimes. If each member of our imaginary community possessed such a right and exercised it by authorizing that all others in the community be placed under the discipline of the devices not to commit offenses against him, then the activation could, on the present view, be completely justified.

The claim that there is such a right is made plausible by considering other, more familiar, self-protective rights that permit us to create serious risks for wrongdoers. First, we may mount appropriately limited, violent self-defense against attacks on our persons and property. Second, we may erect barriers, such as difficult-to-scale fences, to prevent such attacks. Third, we may arrange that an automatic cost *precede* or *accompany* the violation of some right, a cost that is not designed to frustrate the violation but rather to provide a strong reason not to attempt it. The one-way tire spikes placed at the exits of private parking areas provide a commonplace example. And fourth, we may confine those who have shown in the crimes they have committed that they cannot be controlled by other strategies of self-protection.

The first, second, and fourth of these familiar rights are, most fundamentally, rights to render or try to render someone incapable of committing or consummating some crime or crimes. In defending ourselves,

we try to frustrate an offense by disabling the offender.[21] The harm we thereby create may be justified as a means of incapacitating the criminal *or* as an unavoidable side effect of an attempt to block the crime. In erecting very high fences or in confining the incorrigible we attempt to render would-be offenders unable to undertake various offenses. The third right (like the right to which we appeal in activating the devices) has, however, a different strategic character. In attaching costs to crimes, we attempt to prevent an offense by giving the would-be offender reason not to undertake it.[22] But both the basic strategies of incapacitation and threat are designed to protect, and both involve, as we have seen, a connection between the commission of a crime and the possibility of a resulting evil.

Each of these familiar four rights of self-protection has the same wide scope, holding not only within civil society but also in a state of nature. Individuals in a state of nature may defend themselves, set up obstructive barriers, establish automatic costs to accompany violations of their rights, and confine those who are incorrigibly dangerous. I also claim that they have the right to which the activators appeal, the right to use devices that promise costs to follow violations of rights. I cannot, of course, prove this. But reflection on certain features of the other rights of self-protection makes this claim plausible. Consider, for example, the form of self-defense in which the evil that results for the attacker is a side effect rather than a means of defense. The moral acceptability of this form of self-defense shows clearly that the evils created by a legitimate strategy of self-protection need not be justified by reference to *their* effects.

The comparison with the right to create threats of evils to precede or accompany an offense provides even more support. Suppose the best fence that someone in a state of nature can erect to block an attack on his life cannot stop some vigorous and agile enemies. He would then, under this right, be permitted to place dangerous spikes at the top of the fence in order to discourage those who could otherwise scale it. These spikes, like the more familiar ones in parking lots, would not stop a would-

21. Philippa Foot points out to me that one might do this by psychological rather than physical means, by falsely saying, for example, "Your son has been killed."

22. The prospect of a vigorous self-defense may also, of course, give a would-be offender such a reason. But the justification for risking injury to an attacker in defending ourselves does not, I am supposing, arise out of our prior right to create the threat of this injury. If one thinks otherwise, then the justificatory structure I allege to be present in the case of punishment is already present in self-defense.

be offender who is willing to accept any cost, but they would provide most would-be offenders with excellent reasons to hold back. But suppose, to take the story one step further, our defender cannot arrange the spikes so that they offer a threat of injury to someone entering his territory but can arrange them so that they clearly offer a threat of injury to an enemy leaving his territory after an attack. And suppose that the latter arrangement would discourage attacks just as effectively as the former. It would, I submit, be very odd to think that he could have the right to build the first kind of fence but not the second.[23] What morally relevant difference could it make to a would-be wrongdoer that the injury whose prospect is designed to discourage him will come earlier or later? In either case, the injury is not there to stop him if he tries to attack but rather to motivate him not to attack. But building the second kind of fence is nothing more than creating an automatic cost to *follow* an offense. It is, in fact, a very primitive kind of m-retaliating device.

If we do indeed have the self-protective right to create the prospect of such costs, then each of us would, prior to the establishment of a public system of protection, have the right to protect himself by activating a suitably programmed personal m-retaliation device like the public devices that, thanks to our combined resources, we now possess. Of course, even in our imaginary situation we do not have such personal devices, but the fact that we would have the moral right to use them bears directly on our present problem. For each of us may now contribute that private self-protective right to the general authorization to activate the public devices we actually possess. But here a new problem arises. For even if everyone, as I am supposing, prefers protection by the devices to protection by other available means, not everyone is likely to prefer protection by the devices *and* the risk of suffering m-punishments to protection by other means *and* the risk of suffering from the use of those means by others. Some people, whom we may call *rejectors*, will surely fear the increased risk of suffering for their own future crimes more than they welcome the increased protection against the crimes of others. But we may safely

23. It would not, in every case, follow from the fact that he could build the second kind of fence instead of the first that he could build a fence that combined the two threats. My view here is that there is a single self-protective right to attach deterrent costs to offenses whether these costs are to precede, accompany, or follow the offense. If this is correct, then it is plausible to think that there must be a common limit to the total costs that can be attached to a given offense. The fence that poses a two-way threat might exceed this limit.

assume that most people, whom we may call *acceptors*, will prefer the practice of m-punishment, all things considered, and would therefore agree to its full implementation.

It is the acceptors who will, in the situation I am imagining, bring about full activation of the devices in a series of partial activations. Their first step will be to activate the devices with a program that places everyone under threat not to violate the rights of any acceptor. In doing this each acceptor will draw on his own right to make private self-protective threats. The acceptors will then proceed to ask each rejector, in turn, to permit them to extend the operation of the devices so that all are placed under threat not to violate *his* rights. Since even the rejectors welcome the increased protection against crime provided by threats of civic m-punishments, each will be willing. For in accepting this offer of protection a (former) rejector incurs no new risk for himself. His risk of being m-punished comes not from his own acceptance but from that by others. Each person's right of self-protection will thus be exercised and everyone will be brought under the discipline of the devices.[24]

It is important to see that this justification for placing rejectors under the threat of m-punishment is not an argument from fairness. It is not that a rejector may be made subject to the threats because he himself wishes to be protected by those same threats as they fall on others. If that were the argument, the rejector could escape the liabilities of the practice by simply withdrawing from its protections. Instead, the situation in a civil society is the same as in a state of nature, where it is clear that one person may make reasonably limited threats of m-retaliation against others quite independently of whether any of them returns the threats. The source of the first person's right to threaten lies in his legitimate interest in safeguarding himself from the possible misconduct of the others rather than in what the others must, in fairness, accept in return for their threats against him. It is for this reason that one cannot gain moral exemption from these threats by renouncing the use of them. (The same thing also holds of self-defense. One does not lose the right to defend onself against a wrongdoer if, strangely, he has renounced the right to defend himself.) So once a rejector sees that, whatever he decides, others

24. While the acceptors would have agreed to the full practice, and so to their own liability to m-punishment, the full practice does not come about by way of any such agreement. As it happens, no one actually agrees to his own full liability to m-punishment. For that liability always arises by way of other people exercising their right of self-protection.

may justifiably subject him to the discipline of the devices, he will see that he has every reason to authorize his own protection.[25]

Our justification for activating the devices is now complete. Because of its special character, it escapes the earlier objections to the familiar deterrent and retributive theories of punishment. It is clear, for instance, that in placing someone under a threat in hope of keeping him from crime, we are not using him. This does not mean, of course, that threatening in the interest of self-protection is never morally objectionable. To threaten someone is to bring a certain kind of force to bear on him (a force that we identify as intimidation when it is not justified), and such force in human affairs is often wrong. But threatening a person so that he will act in certain ways and using him so that others shall act in certain

25. Of course the rejectors might agree among themselves to refuse the protection of the devices and to form instead a partial state of nature within the community, in which their relations with each other would be reciprocally disciplined by threats of retaliation and their relations with acceptors would be governed by their own threats of retaliation and the acceptors' threats of m-punishment. But there are reasons to doubt that any agreement to execute such a limited escape from the devices would be clearly enough in their interest to be stable. One gain from such an agreement would be the reduced risk involved in violating the rights of other rejectors (since retaliation is less sure than m-punishment). One loss would be an increased risk of misplaced retaliation by other rejectors. (Acceptors are less vulnerable to this risk since rejectors know that the devices will punish all improper retaliations.) But the gain is of doubtful importance. For to the extent that the rejectors' choice to forego protection by the devices indicates that they have little to protect, they may be unattractive as targets of criminal opportunity. (While disadvantaged criminal rejectors may, for reasons of convenience, prey largely on other disadvantaged rejectors, they may stand to gain relatively little real advantage from this.) And the loss is significant. For to the extent that the rejectors' choice indicates that they are indifferent to the moral order, the risk of misplaced retaliation by them is indeed frightening. Thus even if the rejectors would prefer no practice of m-punishment at all, they will have reasons to prefer a total practice to a partial practice of the type in question. Moreover, whenever any party to such an agreement broke faith and went over to the acceptors there would probably be, if the above is correct, even more reason for others to follow. (And such defection could not rationally be prevented by threatening reprisal or even by threatening specially dreadful retaliations for any crimes that defectors might commit against the remaining rejectors. For reprisals or improperly severe retaliations against acceptors by rejectors are themselves crimes that the devices will m-punish.) Of course, if the original acceptors could rightly, as I think they might, make it impossible for the rejectors to communicate, for instance, by surprising each of them with a sudden, irrevocable, and required choice, then the rejectors, assuming that an agreement to reject the practice would be in their collective interest, would be in a prisoner's dilemma. For each would see that no matter what the others did, he would be best off choosing more rather than less protection. Assuming that a fully general practice of m-punishment would be socially beneficial overall, this is a case in which a prisoner's dilemma would work for rather than against the community as a whole!

ways involve quite different moral relations to his will. That a threat is designed to make the threatened party behave as he morally should is a fact that gives it, if not full justification, at least some moral support. However, the fact that an injury to someone helps keep others in line is almost nothing in its moral favor. This means that while a right to punish a criminal in order to deter others cannot be basic (but must, when it exists, derive from some more fundamental right), a right to compel potential criminals to respect one's rights could be basic.[26]

Nor is it the case that we must appeal to forfeiture in order to explain our rights of self-protection. In fact, such an appeal would be subject to the earlier objections. We get a better explanation of these rights if we focus on *actions* and the protections that morality may assign to or withhold from them rather than on *agents* and the general rights they may keep or forfeit. Innocent actions that do not menace others are *morally protected.* This protection consists in the fact that we may not in general attempt to prevent them coercively or frustrate them violently. Violations of important moral rights are, on the other hand, *morally exposed.* That is, we may try to prevent or frustrate them by means that would, in other contexts, violate their agents' rights. That morality should withhold some protection from some seriously wrong actions is easily understood. For these are the very acts that, morally speaking, should not take place. And, to draw a final contrast with retributivism, the explanation of why rights must be contoured so as to permit threats, namely, the appeal to the need to protect ourselves from crime, has an obviousness and compelling clarity missing in retributivists' accounts of the right to punish. That morality should expose would-be wrongdoers to threats in order to prevent wrongdoing is easier to understand than that morality should expose actual wrongdoers to retribution.

Before concluding this examination of m-punishment, we must briefly consider the upper limits on the severity of the m-punishment that may justifiably be threatened for a given type of crime.[27] That there are such

26. It will become clear in Section III that I think it *can* be morally permissible to punish someone with the intention of deterring others, so long as the right to punish is independently secured.

27. In "The Doomsday Machine" Alexander claims that we may threaten those who are competent and free with *any* penalty for any violation of our rights. He argues from an alleged lack of constraints on the self-protective right to construct dangerous barriers such as moats and electric fences. But I think there are substantial constraints even here, constraints that may easily be obscured by the fact that these barriers are generally created

limits and that they are a proper part of the morality of self-protection can be seen by looking at other self-protective rights. The moral right to defend our property does not permit us to kill a burglar whom we know intends us no physical harm. Nor may we erect an extremely dangerous barrier to prevent harmless trespassing. These examples show that the morality of self-protection contains its own rough standard of proportionality. While we may attempt to prevent more serious crimes by creating risks of greater evils, some evils are too great for some crimes. This idea of proportionality is *not* tied to the ideas of retribution and desert. No retributivist claims that the right of self-defense is a right of retribution, but no retributivist can plausibly deny that the right of self-defense is governed by some requirement of proportionality.

The theory of these limits is complex and difficult, and I can here make only some general and tentative suggestions. Using self-defense as a guide, it seems that we do not have to justify particular self-protective threats by any hard and fast criterion of expected general utility. Someone defending himself against an attacker is not burdened by the need to justify the degree of danger that his defense creates by reference to its chance of success. He is entitled to defend himself in ways that put his attacker at risk of evils that are intuitively proportionate to the intended offense even if there is very little chance that the defense will succeed. The same is surely true of self-protective threatening. A penalty cannot be ruled out simply because the threat of it creates more danger for potential wrongdoers than protection for potential victims of crime.

It also seems clear that self-protective threats are not subject to the

---

to prevent a variety of crimes ranging from relatively minor intrusions to very serious assaults. With such a wide range of protection in mind we tend to allow the barrier to create dangers that might be unacceptable if it were used to prevent only the less serious crimes. Thus it would be more defensible to put a very dangerous electric fence around one's house than around a vacant piece of land from which one wished to discourage poaching. And in the latter case, I think it would be seriously wrong to erect such a fence. One should not be confused here by the fact that we may not be obligated to *remove* a barrier that is no more than appropriately dangerous for typical wrongdoers when we realize that some particular wrongdoer may have a special liability to be hurt by it. That is, after all, a special case. Alexander is scandalized by the thought that we might be able to frustrate or deter a crime but not be morally permitted to do so. One can understand this reaction but nevertheless feel certain that such cases are frequent. It must be remembered, however, that if a wrongdoer has proved himself ready to brave all obstacles that we may properly place in the path of his crimes, we may call upon the right of preventive confinement. But we should not confuse this right with the right to carry out threats.

retributivist's standard of equivalence—that the degree of the threatened evil must equal, or at least not exceed, the degree of the evil created by (or intended in) the offense. For if, in a single self-protective response to a possible crime, we may not create an evil for the would-be wrongdoer exceeding that present or intended in the offense, it is hard to see why the *sum* of our self-protective responses to that crime should not be governed by the same limit. But, intuitively, there is no such overall limit, and its adoption would not seem advisable. Self-protective threats of m-punishments will be our defenses of first resort, serving to keep contemplated offenses from ever eventuating. Their capacity to play this role would be considerably diminished if potential criminals knew that any injury they might receive from a victim's self-defense would reduce their m-punishment. Indeed, such an arrangement would sometimes encourage criminals to persevere when they might otherwise be stopped by fear of a vigorous defense. And in some instances it would mean that criminals could not be penalized at all. Consider, for example, a case in which one breaks the arm of an assailant in an unsuccessful attempt to prevent his breaking one's own arm. It seems absurd to think that we must program the devices to withhold m-punishment in such a case.

It also seems absurd to suppose that strict equivalence sets the limit for any individual self-protective response. We may certainly risk breaking both arms of an assailant to keep him from breaking one of ours. And we may threaten m-punishments that are, by ordinary preference rankings, worse than the evils typically inflicted by the crimes they address.[28] There is, however, one consideration that suggests that the limits on what may be threatened must often be set somewhat lower than the limits on what may be done in self-defense. When a threat is made we cannot be sure that this is our last chance of self-protection. But when we are forced to defend ourselves, it is almost always because our other options have run out.

The aim of self-protection does not, however, provide a carte blanche. Self-defense does not, as I have noted, justify any degree of violence against any attack. And some form of proportionality must surely also be observed in making threats of m-punishment. But it is an important and interesting question just why this should be so. Assuming that a potential

28. This is true for serious crimes that because of standard insurance typically result in negligible net loss to the victim and crimes, such as some violations of privacy, that typically result in no harm at all.

criminal does not have to violate our rights, why must we take care not to threaten him with too much in order to deter him?

Morality, I have claimed, exposes wrongs so that they may be prevented or frustrated. It therefore designs variances in some of our rights so that these rights will not interfere with a range of defensive strategies. The question before us is why our rights do not give way so completely that any defense or any threat may be directed against any offense. The answer, I think, is that they retain some force in order to protect those aspects of ourselves and our lives that go beyond any situation in which we choose to commit a crime. Someone who disregards a serious penal threat jeopardizes not only himself and his interests as they are then, but also himself and his interests as they have been and will be. In imposing limits on the dangers that may be placed in his path, morality refuses, in effect, to regard him at the time of a criminal choice as a fully competent disposer of the whole of himself and his life.[29] We may wonder whether morality would extend this protection to beings who were fully rational and totally consistent over time. But it is surely appropriate for us. For human criminals, like the rest of us, have interests and psychological identities that vastly exceed what they can see and defend in a single part of their lives. Morality requires some respect and protection for these larger components of a criminal's identity and good even while it permits us to protect ourselves against him. The result of these conflicting moral pressures is, as one would expect, a compromise.

This compromise naturally results in an upper bound to what may be threatened for a given crime, a limit that wisely allows more serious punishments to be threatened for more serious crimes. Such a constraint sets a limit to the worst thing that may happen to a wrongdoer as a direct effect of the threat against him.[30] A limit of this form can also be defended from the point of view of the comparative importance of the various rights involved.[31] If one is trying to protect oneself from having one's pocket

29. This is, in one sense, paternalistic. But not in the way in which paternalism is usually thought to be objectionable. Objectionable paternalism prohibits people from doing what they may wish to do on the ground that it may be bad for them, and so causes complaints from those who are protected. The present constraints, however, raise objections not from those whom they protect but from those whose protection they limit.

30. By a direct effect I mean one that falls within the direct intention expressed in the threat. If the threat is a threat of death, then death can be a direct effect. If, however, the threat is of a certain term of imprisonment, and such imprisonment quite accidentally happens to cause death, the death is not a direct effect of the threat.

31. See Hurka's "Rights and Capital Punishment," p. 652.

picked one should simply not have the life of the potential pickpocket at one's disposal as material from which to fashion threats. A credible threat of death for such a crime would be a grave moral indignity which even the certainty of deterrence would not diminish.

The ceilings on what may be threatened for different categories of crime must also vary to some extent with several factors other than the seriousness of the offense. First, it is plausible to think that the ceiling may be raised for persons who are especially dangerous, people who have shown themselves ready to commit serious crimes.[32] Second, the ceiling may be raised for crimes that are especially prevalent or are threatening to become so. Here the limit on the threatened evil may be raised generally and not only for people who have demonstrated their particular dangerousness. But this factor should be allowed much less influence than the first. For it is natural to suppose that circumstances largely beyond one's control should not significantly increase one's penal liabilities.[33] Third, the ceiling may be raised for crimes whose detection rates are especially low, at least if this promises a noticeable gain in prevention. But this too would have a limited impact for the reason just mentioned. (This third factor, it should be noted, creates a special theoretical problem that will be important later. In effect, the more severe m-punishments introduced by this consideration are justified only because the guilty are less likely to suffer them. And I believe this means that what is justified in such a case is not the straightforward threat of an m-punishment but the threat of a certain [no doubt vaguely specified] *probability* of receiving an m-punishment. In real practices of punishment the analogous threats are, I shall claim, justified only as threats to *try* to punish.)

II

In imagining our new community and its amazing devices we have, in effect, examined a particular type of protective social practice, the practice

32. Of course, at the moment someone commits a serious crime he is undeniably dangerous. Dangerousness must therefore refer to the general disposition of a person toward the type of crime in question during some fairly long period of time. Taken in this sense, a person may commit a crime without having been dangerous.

33. But to allow this factor to have some influence is not to use the additional jeopardy to a particular potential criminal as a means of deterring others. The additional strength of the threat against him simply addresses the fact that he is a member of the community and that members of the community have, in general, shown an alarming tendency toward the crime in question.

of m-punishment. One of its features, the fact that no persons have to carry out the threats, clearly distinguishes it from punishment. And the fact that it is fully grounded in rights of self-protection might appear to mark another difference. My task in the rest of this article is to show that neither of these features is incompatible with the thesis that acceptable practices of m-punishment and acceptable practices of punishment have, *au fond*, the same moral nature. In this section I shall address the second feature, arguing that the penalties of any intuitively justified practice of punishment would be what we should expect if the moral point of the practice were self-protection. Some terminology will be helpful here. Let's say that a possible practice of punishment and a possible practice of m-punishment are *counterparts* if both threaten penalties of just the same severity for just the same crimes.[34] The thesis that I now wish to defend is this: Every intuitively justified practice of punishment has as its counterpart a practice of m-punishment justified by rights of self-protection, and vice versa. I shall call this the thesis of the functional moral equivalence of counterpart practices (or, for short, the thesis of functional equivalence). This thesis is composed of two claims: first, that exactly the same offenses are properly penalizable in each practice and, second, that all offenses properly penalizable in both practices are penalizable to exactly the same degree in each.

We may take up the somewhat less difficult question of degree of penalization first. It is hard to see how punishments or m-punishments could ever, in being too mild, violate the rights of those who come under threat of them. Our question therefore becomes whether some properly penalizable crime might be subject to a justified threat of a certain punishment even though the counterpart threat of an equally severe m-punishment would be too harsh, or vice versa. Now it seems hard to imagine that a punishment for a given type of crime might be acceptable but the counterpart m-punishment too severe. If a crime is serious enough to be punishable with a severe penalty it must be a very unwelcome violation of rights and therefore subject, under the right of self-protection, to the threat of a severe m-punishment.

However, it might seem that justified m-punishments could be *more* severe than justified punishments for the same crime. For we have seen that the right of self-protective threatening is not subject to retributory

34. I shall also speak of similarly positioned threats and penalties in counterpart practices as counterpart threats and counterpart penalties.

equivalence as a limit on what may be threatened. But, so far as I can see, our ordinary intuitions about particular punishments reject this limit just as decisively. At least this is true if retributory equivalence is determined by anything like usual preference rankings. Consider, for example, our typical legal penalties for crimes against property. Surely the average person, even the average thief, would prefer to have his car stolen than to be confined for a month or two. And the same disregard for retributory equivalence is present in many common punishments for assault and molestation. Not very many people would prefer spending six months in a typical American jail to receiving a serious beating that left no long-term disability. Yet this sentence would not seem overly harsh as punishment for such a crime.[35]

Alan Goldman, who remarks on these disparities, finds our intuitions in these cases are paradoxical because he holds that a theory of forefeiture is the most plausible account of the right to punish.[36] In rejecting that account, however, the present theory of punishment rejects the paradox. When considered in light of retributive principles, these widely accepted punishments can, it is true, seem absurdly high. But if our intuitions are to provide any kind of useful touchstone we must not ignore them when they operate most independently of theoretical preconception. And I can see nothing in our actual working intuitions about the upper limits on degrees of punishment that is hostile to the idea that punishments may be set as high as m-punishments. Moreover, each of the special factors that can properly influence the severity of self-protective threats seems equally capable of influencing our feelings about punishment. Consider, for example, the higher penalties we often assign to repeat offenses. Past conviction operates here as a criterion of dangerousness with respect to the type of crime in question. It is also true that we sometimes feel justified in threatening somewhat greater punishments for crimes that are especially prevalent in the community as well as for crimes that pose especially difficult problems of detection and conviction.

Having looked briefly at the question of degree, we must turn to the

35. A disproportion between harm done in the commission of a crime and harm received in punishment can hardly be avoided in punishments for attempted but unsuccessful crimes. Here retributory equivalence must refer, I suppose, to the harm intended by the criminal.

36. "The Paradox of Punishment," *Philosophy & Public Affairs* 9, no. 1 (Fall 1979): 42–58.

more difficult question whether counterpart practices would justifiably threaten penalties for just the same offenses. At first, it seems clear that wherever we properly threaten punishment the counterpart threat of an m-punishment would be equally justified on purely self-protective grounds. This is true even for threats against attempted crimes. If we did not bring substantial attempts under threat of m-punishment, would-be criminals would know that they could, without risk, always place themselves in an advantageous position from which to decide whether or not to risk committing penalizable crimes. They could start the project and then decide to abort it if the risk of proceeding seemed too great, or to complete the project (under the favorable terms secured by their preparation) if the risk seemed low. Not to place some attempted crimes under threat of penalty would be very dangerous.[37]

A more serious challenge to functional equivalence is presented by people who give evidence that they are incited by the prospect of penalties to commit the very crimes to which the penalties attach. It might seem that such people would, if not compulsive or incompetent, be punishable even though it would not be sensible to place them under self-protective threats of m-punishment. In one kind of case (certainly the most familiar), we may know that a person is *sometimes* led toward crime by the prospect of a penalty that usually deters him. It is plausible to think that we may let ordinary general threats of m-punishment stand against such people, even against those particular crimes that they may commit because of the threats. For were we to exempt such crimes, knowledge of the exemption might encourage these people to indulge their more ordinary criminal motives in the hope of seeming to have acted from the

37. The punishment of unsuccessful attempts to commit crimes presents theoretical difficulties for more than one conception of punishment. The class of attempted burglaries (whether completed or not) is wider than the class of burglaries (since all burglaries have been attempted but not all attempted burglaries succeed), and therefore we have more to fear from a burglary than from an attempted one. This suggests that the ceiling on the self-protective threat against attempted burglary should be set somewhat lower than the ceiling on the threat against completed burglary. On the way of looking at the matter most congenial to my theory, we do not make any threat specifically directed against attempted *but unsuccessful* burglary, since that class of actions is not particularly dangerous. Rather we make a threat against attempted burglary (whether successful or not) but stipulate that, in cases where an attempt succeeds, the force of the threat is preempted by the force of the threat against actual burglary. As Bentham pointed out, it would be a serious defensive error to make the penalties for attempted crimes as severe as those for completed crimes, since that would eliminate an important incentive to abort crimes under way.

special motive. There seems nothing unjust in increasing our overall protection against someone who is free and able to avoid committing crimes by allowing threats to stand against him on the rare occasions when they do more harm than good. In another (and certainly rarer) kind of case, we may know that a person is so frequently and strongly incited by the prospect of penalties that he would be less dangerous overall were he exempted from the usual threats.[38] We might, of course, be inclined to keep him under them for fear of the effects on others of exempting him. For if we did exempt him, some people who do not suffer from his condition might commit crimes in the hope of appearing to suffer from it. But to refuse for this reason to exempt the known victims of this kind of irrationality would be morally questionable. It would seem to be a matter of using them to gain protection against others.

I am, therefore, inclined to think that such people may not be brought under self-protective threats of m-punishment. And this means that there is, in theory if not in practice, a class of free, but extremely irrational, wrongful acts that may not be m-punished. But how do such offenses stand with regard to punishment? It is not, I think, counterintuitive to judge that the people who commit them are irrational and abnormal in a way that throws doubt on their fitness for inclusion in a genuine practice of punishment. To present them with the prospect of punishment would, in effect, be to *invite* them to commit crimes. A certain kind of retributivist will, of course, disagree with this exemption, finding in this kind of perversely undeterrable crime need for the most stringent, and therefore the most criminally exciting, punishments. But here, I cannot help thinking, retributivism shows itself in a disadvantageous light.

We must now turn to the question whether every m-punishable crime is also punishable. Here different problems arise. The cases that most

38. Note that it might be sensible to place such a person under threat of m-punishment for any criminal act that he can be determined to have committed from more ordinary criminal motives. And if he were sufficiently alert and had enough self-control, these special threats might give us all the protection we could get in the normal case. For the moment that he became aware that he might be about to commit a crime in some perverse reaction to being threatened, he would remember that the special threats do not, then and there, apply to him. The reaction would therefore subside, and he would be brought into a more rational relation to the special threats that stand against him. Substituting these special threats would not make sense, however, if he could not remember their special character or if he could not control his emotional reactions once they had begun. But in that case, he is under such serious psychological handicaps that we may surely doubt that he is fit for punishment.

call for examination are those in which real punishment seems illegitimate despite the fact that an objective violation of rights has occurred. The question we must ask is whether these cases would also be excluded from a just self-protective practice of m-punishment. To begin, we may consider punishment of innocent third parties, for example, punishing parents for crimes committed by their children. Such a form of punishment is certainly ruled out by ordinary moral intuition. But threatening third parties might be a very effective means of self-protection. Nevertheless, I think we can see why these threats would be morally illegitimate considered strictly under rights of self-protection. For whether or not they succeed in deterring a given crime is not ultimately in the hands of the party who is in risk of receiving the penalty. And, no matter what the gain in protection, it is manifestly unjust to threaten to inflict an evil on someone when it is not up to him to do that which will prevent it.[39]

It is also important to consider cases in which punishing would not be justified because of the criminal's special mental condition at the time of the crime. We may first consider the already mentioned case of compulsion.[40] It might seem that there could be no self-protective point to placing compulsives under threat. But this is incorrect. If genuine com-

39. That this third-party constraint holds properly of the morality of self-protection (and not just of the morality of retribution) can be seen in the case of self-defense. We can construct imaginary examples in which a wrongful attack by one person could be physically frustrated by means of violent reactions directed against another person who is no party to the attack. Imagine a pair of Siamese twins, A and B, joined so that both will be seriously injured or die if B is shot but so that B will not die or be seriously injured if only A is shot (if A is killed B may be surgically separated and saved). Suppose that A assaults you with the intention to do you grievous bodily harm (dragging the reluctant and vainly struggling B along). You have a gun, but are physically prevented from aiming it at A. You can, however, injure or kill A by shooting B. But surely, even though you would be within your rights to shoot A, the attacker, you cannot shoot the innocent B even to protect your life. This is not because injury or death is something that B does not deserve. If you were able to shoot and injure A his injuries would not, I think, be counted by the theory of retribution as part of his just deserts. These would come only later in his proper punishment. You may not shoot the innocent B because you are not defending yourself against *him*. This very fundamental constraint on activities of self-protection is primarily a matter of what may be in itself intended or done as a means. We must not let it be obscured by overattention to cases in which self-defense puts nonattackers at risk incidentally. Third-party threats would, of course, make direct rather than incidental use of a danger that is not in one's power to prevent.

40. Perhaps compulsion is a matter of degree. If so, then when there is at least some freedom there may also be room for limited threats and hence for limited penalties. Note also that threatening certain people might actually enable them to break the grip of what would otherwise have been their compulsions.

pulsives were excluded, some people who are not true compulsives would be encouraged to commit crimes in the hope of seeming to have been compelled. And since neither we nor the devices can ever be absolutely certain that someone acted wholly by compulsion in committing a crime, even those who give every indication of having been compelled may actually be faking. We therefore have deterrent reasons for refusing to exempt genuine compulsives from our threats. And the thought that we might properly do this is encouraged by a possible analogy with self-defense. We are not obligated to worry about the chance that a defense may be unable to frustrate a crime. Indeed, we are entitled to defend ourselves in ways that can harm an attacker even if we are virtually certain that we cannot succeed because of, say, the attacker's strength. Why then should we worry about the fact that our threats in some cases probably cannot succeed because of someone's compulsion?

The difference is this: Self-defense against an actual criminal is justified as a way of disabling him, while threats are justified as a way of giving a potential criminal reasons. Defending oneself is therefore an activity in which the attacker is simply acted upon. In threatening, however, one assigns a morally essential active role to the threatened party. He is to consider the reasons he is given by the threat and is, if all goes according to plan, to refrain from certain criminal choices at least partly for those reasons. It is one thing to cast someone in this role who (we are certain) will ignore the reasons. But it is quite another to assign the role to someone who cannot be influenced by them. That would be unjust. Nothing like this injustice is present, however, in the case of self-defense that cannot succeed.

The same argument applies, of course, to any mental state in which a person is unable to take account of his reasons for action, for example, hysteria, extreme depression, and a variety of mental illnesses. A related but somewhat different reason applies to people who do harm unwittingly without thereby violating any duty of precaution. Someone who acts in this way is simply not in a position to bring a threat of m-punishment into his deliberation about what he sees himself as doing. Placing such choices under threat of m-punishment would, again, be unjust. It would assign a role to the threatened person that he will be unable to play given what he knows.[41]

41. Acceptable threats to hold someone strictly liable for a type of proscribed action are, I think, best conceived as threats that warn people to take extreme precaution not to do

Suppose, however, the problem that raises the question of punishability is not lack of freedom or relevant factual knowledge. Suppose someone knowingly commits a crime who is free in the sense of being able to engage in and act on his practical deliberations but is unable to understand or appreciate the moral order behind the penal code. Suppose, for example, that we were to discover the existence of a genetic amoralism, a condition that deprives its victim of any moral sensibility or internal moral motivation but does not affect any other cognitive or deliberative faculty. Some genetically amoral people are, we discover, good citizens in whom any antisocial motives are held in check by the nonmoral civilizing motives that affect us all: desire for the esteem of others, fear of disgrace, and especially fear of civic sanctions. Others, perhaps a significant proportion, are criminals. Now, surely we could threaten those genetically amoral people who are free, clearheaded and concerned for their own well-being with m-punishment in an attempt to deter them from crime. But since they are not, in some sense, morally responsible for what they do, could we punish them? Here again I think it is important to consult our moral intuitions in practice rather than when they have passed through the filter of retributive theory. If we do, I think we shall conclude that whether someone is to blame for his own amorality or immorality is, by itself, irrelevant to our actual decisions when and when not to punish him. We routinely punish and, I think, rightly punish sociopathic criminals whom we have absolutely no empirically respectable reason to blame for their conditions. What matters to us is whether

---

the action, even unintentionally. If someone has taken every extreme precaution that can reasonably be required, he may not, on my view, be penalized. The strictness of strict liability cannot properly be strictness in principle. Holding someone subject to threats for forms of behavior that he, through his own fault, does not know to be proscribed is, I think, a special form of strict liability. (Any adult would, of course, have to be mentally defective not to know that the kinds of actions we have been discussing are crimes.) Each of us is, in effect, under threat to take care to learn what is proscribed by the penal code before we act in a manner to which the code may address itself, and the penalty for a failure to take this care, when this leads to the commission of a crime, is the very penalty set for that crime. On this view, when someone is properly penalized for a law of which he was ignorant, the threat with reference to which he is penalized must involve the general (and generally understood) warning to take care to inform oneself about the contents of the penal code. And the evil referred to in this threat is not a penalty, in the usual sense, but rather the prospect of receiving one of a number of different penalties in case the failure to inform oneself leads one to commit one of a number of different crimes. Such threats are special in lacking a full description of the threatened danger. But, since one can always find out the precise character of the danger before one runs afoul of it, they can be fair.

they clearly understood the threats against them and were capable of being deterred by them.[42]

But what about the more extreme case of people who are free in the way we have been considering but whose moral *and* nonmoral sense of reality is not as it should be? Take someone who believes that God tells him to kill as many people as possible. Of course, we sometimes suppose that people who suffer from such massive delusions are in the grip of these delusions in a way that undermines their freedom. But perhaps this is not always true. We must therefore consider possible cases in which people who are disturbed in their thinking but are nevertheless free and able to deliberate, commit crimes. Now we certainly feel that punishment ought to be ruled out in many such cases. And the question is whether the practice of m-punishment would treat these cases in the same way. There is, I think, one reason to think it would. For even if a madman can deliberate, he may not be able to grasp the danger posed by threats of m-punishment; or while he may in some sense grasp the danger, he may be unable to give it proper weight in his deliberation. The rationale for not including such people under threats is therefore the same as that for excluding compulsives. It would be unjust to create dangers for them that they cannot escape or cannot have a reasonable chance of escaping.

But perhaps not all mad criminals are unable to give threats of m-punishment enough weight in their deliberations. Perhaps there are some who retain enough hold on reality to appreciate the full force of the threats. Indeed I suspect that we are confronted with just such crazy but deterrable people in increasing numbers—terrorists and fanatics who act in the name of insane causes but who seem, since they take considerable trouble not to be caught, capable of being influenced by threats. Such people could certainly be placed under threat of m-punishment. But it is equally true, I think, that they may be threatened with punishment properly so called. And I am sure that the thought that they are punishable accords with the ordinary judgments of most of us. It is undeniably true that there is a sense in which such people are often not

42. Our own legal systems make knowing right from wrong a condition of punishability in at least certain kinds of cases. But in practice I think that this condition cannot require more than that one be well aware of the contents of the moral code of one's community and the character of its major moral distinctions. And this is a kind of knowledge that genetically amoral and other disturbed persons could fully possess.

to be blamed for what they do. They are not like those of us who commit crimes from familiar and contemptible motives of greed or lust, and they may be no more responsible for their disturbed outlook than genetically amoral persons are responsible for their lack of genuine moral motivation. But there is a sense in which they may be held responsible for the real crimes they commit. For they commit them freely and deliberately in full knowledge that they are under threat designed to deter them. This sense of responsibility is usually enough to satisfy our everyday sense that punishment is in order, and, if I am right, it also ought to be enough for us in theory.

Two observations may help make the thesis of functional moral equivalence more plausible in these cases of mentally disturbed criminals. First, the fact that a disturbed person is at various times deterred by threat of m-punishment does not entail that he is, in the relevant sense, deterrable at the time he commits a particular crime. Justice requires that a threat apply to a criminal choice only if *in making that choice* (or the choices that lead to it) the criminal is able to understand the threat that applies to him, is able to appreciate the threatened penalty as something unwelcome, and is able to avoid the crime. Second, we have imagined that the devices are programmed to mete out m-punishments for any crime that was properly subject to our self-protective threats. But there can be genuinely humane (and therefore moral) reasons not to punish a crime, especially when the offender is mentally disturbed, that do not call into question our right to punish. If we wish to get a fair comparison of the two practices, we must either imagine ourselves as punishing whenever we see ourselves as having a strict moral right to do so or, preferably, we must imagine the devices to be programmed with principles of humanity as further constraints on their operation.

III

The thesis of functional moral equivalence should, to the extent that I have succeeded in making it out, incline us to take seriously the idea that the moral essence of a just practice of punishment and that of its counterpart practice of m-punishment are the same, that both are systems of deterrent threats fully justified by rights of self-protection. But a difficulty remains. For real punishment involves not only creating threats but also carrying them out and therefore raises questions that do not

arise in the case of m-punishment. While the dangers to potential wrongdoers may be no greater under a practice of punishment, their realization will require real persons to perform various real actions all of which will clearly stand in need of some kind of moral justification. And it may seem that no appeal to our right to protect ourselves from possible crimes could serve to establish a right to do anything about those crimes once they had become actual. But the problem is, in a way, even worse. For the right to which we appealed when activating the m-punishing devices was the right to attach *automatic* costs to crimes. But in the case of punishment we need to appeal to a right to attach costs that will have to be imposed by human agents. Thus we seem forced back upon a particularly acute version of the difficulty with which we began.

But while it is true that the move from threats of m-punishment to those of punishment generates these new and difficult philosophical problems, it is, in my view, a mistake to assume that only an *independent* account of the right to punish can solve them. For this is to assume that the right to establish the threat of punishment is posterior in the order of explanation to the right to punish. But while very natural, this assumption is, I think, mistaken. In my view, the right to establish the real threat of punishment is the moral *ground* of the right to punish. I shall presently try to defend this hypothesis. But first we should briefly consider how the hypothesis would, when added to our previous results, enable us to reach the conclusion that practices of punishment and m-punishment rest on the same moral foundation.

To say that the right to establish a genuine threat is prior to the right to punish (or that the former right grounds the latter) is to make two claims: first, that the right to set up the threat can be established without first raising the question of the right to punish and, second, that the right to the threat implies the right to punish. According to the first claim, a case that prescinds from any consideration of how one will later be justified in punishing and concentrates exclusively on what is to be said for and against the creation of the real prospect of punishment for crime (that is, the real likelihood that a criminal will be punished) can be sufficient to establish the right to the threat of punishment. If this claim is true, then in our moral deliberations about setting up a practice of punishment we may regard the creation of the threat *as if* it amounted to causally determining our wills so that we would in fact try to punish

crimes and would do so without raising any further question of right.[43] But this is to claim that the right to set up the threat of punishment may be treated as if it were the right to threaten that which will come about automatically, that is, as a causal consequence not subject to a certain kind of further moral scrutiny.

The first claim therefore implies that the right to establish threats of m-punishment and the right to establish counterpart threats of punishment are on the same moral footing, that the right to attach automatic costs must generalize into a right to attach costs that are either automatically or personally imposed. For apart from the fact that the threat of punishment is the threat to *do* something (the fact that we are to set aside), the morally relevant structure of the situations in which we establish the counterpart practices is the same. In both there are holders of rights who wish to protect themselves from potential violators of these rights, and in both there is the possibility of creating conditional dangers that will tend to deter crime. And according to the second claim (that the right to the threat implies the right to punish), considerations that suffice to establish our right to the threat of punishment will also suffice to establish the right to punish when the time comes. Since we are now justified in creating the real prospect of punishment we will later be justified in punishing. The thesis of the explanatory priority of the right to create the threat of punishment thus means that a practice of real punishment, both at the time it is established and later, has the same basis in moral rights as its counterpart practice of m-punishment.

But why should we believe that the right to establish the threat is prior? The ultimate plausibility of the hypothesis lies in the fact that it gives a more satisfying account of the right to punish than any alternative. The best defense here, as in the case of other highly theoretical moral claims, is an argument to the best explanation. I shall not, therefore, attempt any kind of proof. There are, however, certain reflections that can make my hypothesis seem dubious (or even incoherent), and it is to some of these that I now turn. One line of thought begins with the way in which establishing real threats of punishment must involve the formation of conditional intentions. In the simplest case an individual threatener must himself form an intention to punish crimes, and in more complex cases

43. That is, not raising any further moral questions about our right to punish kinds of crime that are properly placed under the threats.

various members of a penal establishment must form various intentions that together could be thought of as a collective intention to punish crimes. Once one sees the role of intention in the creation of a threat of punishment, one will be reminded that, in standard cases, the moral justification for forming an intention is dependent upon the justification one anticipates having later for doing the thing intended. And, generalizing on these standard cases, it will seem that the justification for establishing a threat could not be prior to the justification for punishing.

That the justification for forming intentions is usually parasitic on the anticipated justification for the thing intended is not surprising. In the vast majority of cases there is nothing of moral interest to assess in the formation of an intention other than the independent moral character of its object. The typical intention has no morally interesting life of its own. What will bear on people's good or ill, respect or violate their rights, is the action intended and not the coming to intend it. In this respect, however, conditional intentions whose expression is designed to deter or induce future action in others form a very special class. This is all the more true when the insincere expression of these intentions is, as in the case of promises and official public threats, morally questionable. In such cases, morality takes an interest not only in what we ultimately do but also in whether or not we form the intention to do it. Moreover, these intentions are embedded in actions, promises, and threats, that clearly have an impact on the good or ill of ourselves and others and therefore have a striking moral character of their own. The conditional intentions to punish or to contribute to a joint undertaking of punishing contained in the setting up of a practice of punishment are not, therefore, typical. It is not plausible to say of them that they have no moral interest apart from that of their objects.

That this is so can be seen by reflecting on the case of sincere promises to pay for future services. On the view that the justifiability of forming an intention always derives from the independent justifiability of its object, we could be justified in sincerely promising to do something only if we could justifiably do that thing independently of having been justified in making the promise. But this is not always the case. We may be morally permitted to do some things only because we *were* morally permitted to promise to do them. For example, the guardian of a ward and his estate would not, typically, have the moral right to disburse the ward's funds to *give* someone something for services rendered the ward in the absence

of any prior agreement, but he surely would have the moral right to *pay* someone for services rendered under an earlier contract justifiably entered into in behalf of the ward's interests. Appeal to the ward's interests here plays a crucial role in accounting for the right to expend the ward's funds. This can be seen by noticing that a promise to pay made by the guardian in the foreknowledge that the services would be rendered *whether or not* there was a prospect of payment might create no moral claim on the ward's funds. If this were so, a plausible explanation of why the guardian may honor a promise made in order to secure the welfare of the ward but may not (with the ward's funds) honor a promise made in indifference or hostility to the interests of the ward is that in the former case, but not in the latter, making the promise was justified in the first place. In such cases the question of primary moral moment arises at the time the promise is made: Can it be justified by the way in which it may be hoped to benefit the ward? And it is important to keep in mind that we are not speaking here of insincere promises. The question is whether the guardian would be justified in making a promise with the full intention to honor it.

The same moral structure is present, I would argue, in the case of threats. What one may sincerely threaten to do in order to avoid certain things is not always determined by what one could do independently of the fact that one had the right to threaten. But suppose it is granted that certain promises and agreements involving the formation of conditional intentions can be independently evaluated in the way indicated. It might, nevertheless, be argued that it is a mistake to think that this can be generalized to the case of threats. For the relevant kind of case arises only when there is a way for the justification of an earlier act to carry forward to a later one. What enables the justification for the guardian's rightful promising to ground the later justification for making payment seems to be the fact that once he has promised, he is *obligated* to pay. Obligation contains permission, and therefore provides a moral medium that can carry an earlier justification forward.

This, in effect, threatens to reduce my view about punishment to absurdity. For, the objection continues, it is absurd to suppose that one could have the right to establish a real threat of punishment but not, *ceteris paribus*, the right to punish. In the case of the guardian's promise the analogous absurdity (that he might have the right to make the promise but not, *ceteris paribus*, the right to honor it) is avoided by the convenient

fact that promises create obligations that carry permission forward. But there is no comparable forward-reaching moral mechanism in the case of the threat. For there is no general obligation on the part of penal authorities to punish every crime they have the right to punish. It must be, the objection concludes, that the moral alignment between threat and punishment is due to the priority of the right to punish rather than the priority of the right to threaten. This is, perhaps, the most serious line of objection to my conception. To escape its force, I must be able to explain how a prior right to establish the threat of a given punishment *could* transfer forward to the right to mete it out. *And I must be able to explain this in a way that never presupposes that the right to punish has been secured first.*

I shall now try to construct just such an explanation. It consists in a series of steps that lead from threatening to punishing such that each step not only implies the next but, if I am right, implies it without presupposing it. The steps are these:

(1) *At* t_1, x cannot object to the fact that we then create a real threat that, if he commits a crime of type C at t_2, we will thereafter try to see to it that he receives a punishment of type P.

(2) *At* t_1, x cannot object to the fact that, if he commits a crime of type C at t_2, we will thereafter try to see to it that he receives a punishment of type P.

(3) *After* t_2, x cannot object to the fact that, if he has committed a crime of type C at t_2, we are trying to see to it that he receives a punishment of type P.

(4) *After* t_2, x cannot object to the fact that, if he has committed a crime of type C at t_2, we are actually seeing to it that he receives a punishment of type P.

(5) *At and for some time after* t_3, x cannot object to the fact that, if he has committed a crime of type C at t_2, we are subjecting him to a punishment of type P.

(6) If x has committed a crime of type C at t_2, then *at and for some time after* t_3, x cannot object to the fact that we are subjecting him to a punishment of type P.

The steps are to be interpreted as follows: "We" refers to all citizens who authorize members of the penal establishment to fulfill their various functions in the case of x's crime. "Seeing to it that x receives a punish-

ment" refers to the complete performance of all these functions, that is, investigating the crime and apprehending, convicting, and fully punishing x for having done it, all carried out in the name of all the authorizing citizens. And "trying to see to it that x receives a punishment" refers to this collective activity insofar as it is begun but is uncertain of completion despite the best efforts of all concerned. (We can, in the intended sense, be trying to see to it that x receives a punishment in the detective work done before x is identified as the criminal.) T_1 is the time at which the threat is created, t_2 the time at which the crime in question is committed, and t_3 the time at which the punishment begins. And at each step, the kind of objection that x lacks is one whose force would show the presence of some moral right and could, therefore, obligate us to see to it that the objectionable state of affairs did not obtain. I shall speak of a state of affairs to which x cannot, in this sense, object as one that is morally acceptable to him.

(1) is, of course, our starting place. It asserts that at t_1 we have a moral right, so far as x is concerned, to create the threat that we will try to see to his punishment if he commits a certain crime. To make the assertion as plausible as possible, we may suppose that x's mental condition makes him a clear candidate for placement under penal threats and that the threatened punishment is appropriate to the crime. To create this particular threat is to shape the present order of things so as to make the conditional (that we will try to punish x if he commits the relevant crime) probable, and to warn x that this has been done. Often, in creating such a threat, we actually succeed in making the conditional *true*. That is, we succeed in affecting the present order of things so that if the threatened party does commit a future crime, we will, in accordance with our present intentions and plans and only because of those intentions and plans, try to see to his punishment.[44] Suppose that in creating the present threat against x at t_1 we actually succeed in making the relevant conditional true in this sense. Assuming that our sincere threat is a morally appropriate self-protective measure, x cannot at t_1 object to our making this conditional as likely as possible. But then he cannot object that in making it as likely as possible we actually make it true. And if he cannot object

44. "Because" here indicates that the attempt to punish would not follow the crime if we had not formed the earlier intentions and plans. It does not, however, express a sense of causality incompatible with human choice. Often we would not in fact make a choice had we not made some earlier one.

to this, he cannot object to the truth of the conditional itself. Moreover, none of these inferences seems to depend on some hidden way in which the implying proposition presupposes the implied proposition. The best explanation of why x cannot at t_1 object to the fact that we will try to see to his later punishment should he commit the crime seems to consist in the fact that he cannot at t_1 object to our creating the self-protective threat that we will do so. (1) therefore implies (2) without presupposing it.

(3) brings the acceptability (unobjectionableness) expressed in (2) forward to a time after the crime has occurred. In both steps the same conditional state of affairs (if x commits the crime at t_2, then after t_2 we try to see to it that he is punished) is said to be morally acceptable to x. In (2), that state of affairs is seen by x from an earlier perspective, and in (3) it is seen by him from a later perspective contemporary with our attempt to punish him.[45] Either step *could*, for all we yet know, be prior in the order of explanation to the other. The earlier acceptability could be derived from the anticipated acceptability of trying to see to x's punishment. Or, as I claim, the latter could be based on the former. In any case, it is certainly tempting to think that one of these steps must explain the other.

But whichever we take to be prior, we have strong reason to reject any suggestion that the two judgments may *differ* in truth value. For if they do come apart in this way, it must be that they express incompatible moral conceptions. Relative to a single morality, a given state of affairs will be morally acceptable to x at all times if it is acceptable to him at any. This is because the notion of having a morally relevant objection that I intend here is *objective*, not a matter of whether x *knows* a good objection but whether there *is*, in principle, one that could be put forward in his behalf. And an objection that could be put forward for x at one time could, in principle, be put forward for him in some form at any other time. If, for example, x can rightly object at t_2 that trying to punish him harms his interests without furthering ours, he could have rightly objected at t_1 that the punishment we then threatened might turn out to be like this. But both we and x then knew that although this unhappy result was a real possibility, it did not provide x with a legitimate objection.[46]

45. (2), therefore, uses the future tense while (3) uses the present tense. But these uses of tense are unessential. Each step could be expressed in a tenseless idiom.

46. The moral structure of our situation here must be distinguished from that of nuclear

Moreover, we should not be mislead by the possibility that new information may arise between t_1 and t_2 that would make the attempt to punish morally objectionable. Even if such a development could not have been predicted in a particular case, provision for it could and, in principle, should have been made. The threat should have been conditionalized so that it would not apply in case the unexpected information did arise. For example, even if we could not have predicted that x would become a kleptomaniac we could and should have restricted our threats against him so that they would not apply in this eventuality. This technique for bringing threats and attempts to punish into moral alignment in no way presupposes the explanatory priority of the one over the other.[47] It is nothing other than an expression of the requirement that (2) and (3) be equivalent. And it is just this required equivalence that explains the possibility of deriving (3) from (2). The important point is that the equivalence holds because the relevant notion of acceptability is governed by a constraint of temporal neutrality and not because an objection to the threat must derive from some prior objection to punishing. This explanation can be understood quite independently of resolving the question which step comes first in the order of explanation.

Let us now consider the inference from (3) to (4). The reference to "trying" in (2) and (3) is, I think, essential if threats of higher penalties can be justified, as I think they can, by reference to unusually low detection rates. In these cases the original threat is acceptable only because we may fail to bring off what we threaten. What x cannot object to at t_1 in these cases is not, in the first instance, the prospect of our actually seeing to it that he receives the specified penalty in full, but rather the prospect of our trying to see to it.[48] This means that we must find a way

---

deterrence. There, in the hope of reducing our chances of being attacked, we threaten to do that which will, among other things, destroy innocent third parties. If such threats could, as some believe, be morally justified, they would be justified despite the good objections of these innocent third parties. Their objections would be overridden by the expectation that the threat will help us avert disaster. Should such a threat fail to deter an attack, our expectation will be proved false and the innocent third parties' objections will remain unopposed by any cogent moral counterargument. The present derivation, resting as it does on the claim that there is *no* good objection to creating the threats, could not even begin in such a case.

47. I have already argued that the morality of self-protective threatening can account for all, or most, intuitively plausible limitations on punishing.

48. I assume here that it would be unintelligible to suppose that a morally unacceptable prospect could be rendered acceptable by lowering the probability of its realization, but that

of bridging the gap between the attempt and the thing attempted. My strategy for this rests on two premises. First, that trying to see to it that x is fully punished is, once the attempt begins to be successful, the very same activity as seeing to it that he is fully punished.[49] And second, that the acceptability of an action or activity under one description entails its acceptability under any other true description. This second premise results directly from the fact that the notion of acceptability contained in my argument is to be understood in an "all things considered" or "*überhaupt*" sense. One has or lacks an objection in this sense to an action (or to a state of affairs containing an action) no matter how the action is described. Seeing to x's punishment will therefore be acceptable to x, all things considered, just in case it is the same activity as trying to see to x's punishment.

My identification of these activities rests on a general claim that trying to do something and actually doing it can be the very same action or activity differently described. I do not claim, of course, that in a protracted but eventually successful attempt to open a door one is from the very start actually opening the door. For such simple actions as this, only the last moment of a successful attempt belongs to the action attempted.[50] But for most activities that include a variety of different actions as parts, actual performance is to be identified with attempted performance from the moment that the attempt begins to succeed. For example, a doctor's attempt to heal someone completely (once it begins to succeed) and his healing that person completely are the same activity. And the same is true, I claim, for a successful attempt to bring someone to full justice and the actual bringing of him to full justice. If our attempt to see to it that x is punished for his crime is, all things considered, acceptable to x, our actually seeing to it must also be acceptable to him.

---

it is fully intelligible that a prospect of someone's attempting something might be acceptable, not because of the low probability that the attempt will occur, but because of the low probability that it will succeed.

49. This is relevant because (1) through (6) can be understood to refer directly to our activities. For example, (3) can be rewritten: "Our *trying* to see to it that x receives a punishment of type P for having committed a crime of type C at t_2 is (after t_2) something to which x cannot object," and (4) can be rewritten: "Our actually *seeing* to it that x receives a punishment of type P. . . ."

50. Even in such a simple case I would deny that a trying can be successful only in virtue of being succeeded by a doing that is no longer a trying. Otherwise there could be no such thing as a trying that succeeds from the very start.

This account of the inference from (3) to (4) is, like that of the inference from (2) to (3), neutral with regard to priorities. Some description of what we are doing in regard to x must, one thinks, be the morally *relevant* one, the one that explains why he cannot object. But when our action is both a trying and a doing, the inference from its acceptability under one description to its acceptability under the other holds whichever description has explanatory priority. That there is no basis for insisting that the acceptability of the doing (the succeeding) must ground the acceptability of the trying can perhaps be seen by drawing an analogy with the logic of permission.[51] You may grant me permission to try to A without granting me explicit permission to A. Indeed, when it is highly unlikely that I can A, permission to try may be all that you are in a logical position to grant. More to the present point, it may be that you are willing and able to grant me permission to try to A if you believe that my chance of succeeding is low, but you would not otherwise be willing or, perhaps, able to grant me permission to A. But if in such a situation you do grant me permission to try to A, there can be no further question about my A-ing having been done with your permission should I succeed. This suggests that it is intelligible to begin with the thought that in certain cases morality grants us a permission to try to see to it that someone receives punishment from which we can *infer* that, should we succeed, we will have acted permissibly.

Thus, both the inferences from (2) to (3) and from (3) to (4) depend on what might be called the logic of the relevant notion of moral acceptability to x and on the fact that the states of affairs judged to be acceptable to x in each pair of statements are, in one way or another, identical. The inference from (4) to (5) is based on the principle that it cannot be acceptable to x that we do a number of things unless, in so acting, it is acceptable to him that we do each of them. The action referred to in (5), punishing x, is a proper part of the activity referred to in (4), seeing to it that x is punished. There is no way in which x could lack an objection to the whole of that activity if he had an objection to this part of it. The acceptability to x of our punishing him is not, however, prior to the acceptability to him of our seeing to it that he is punished. For the latter involves bringing x to trial and convicting him, and x may rightly object to any punishment not embedded within such an acceptable whole.

51. Suggested by Rogers Albritton.

Only the inference from (5) to (6) remains to be considered. Suppose we ascribe to (5) an underlying form in which what x cannot object to is the conditional proposition as a whole. Then the inference to (6) will call upon a modal principle similar to one found in the logic of possibility. That acceptability should be governed by such a principle does not seem odd. For it is very like a form of permissibility, permissibility as seen from the point of view of a person acted upon, and permissibility can be understood as possibility in a normative system. On the other hand, if we ascribe to (5) an underlying form in which under a specified condition (x's committing a crime) x lacks an objection to a nonconditional proposition (that we subject him to a punishment), then the inference to (6) requires nothing but *modus ponens*.

If this account of the relation between these steps is correct, it follows that if a threat of punishment could be fully justified by the rights of self-protection that justify a threat of m-punishment, the force of these rights would carry forward to the act of punishing. If the urgency of self-protection makes moral room for threats it also makes moral room for punishment. But this means that a justified practice of punishment *can* be intelligibly conceived to have the same moral essence as its counterpart practice of m-punishment. It is possible to hold that punishing a criminal for a crime does not violate his rights *because* subjecting him to the threat of punishment for such a crime did not violate his rights in the first place.

I now wish to consider a possible criticism, the reply to which will help bring out an important feature of this conception. The objection is directed against the very idea that the later acceptability of an action can derive from the earlier acceptability of its prospect. It starts with the initially plausible looking assumption that if the explanatory structure I invoke to explain punishment is valid, it should also apply in purely prudential situations. Consider the following fascinating case invented by Gregory Kavka.[52] An eccentric millionaire offers N a fortune to form the intention to drink a toxin that will make him feel rather ill. N would, quite sensibly, be glad to accept such a temporary unpleasantness in order to get the fortune, but that is not what he has been asked to do. He is offered the fortune as a return not for the action itself but for forming the *present* intention to perform it at a later time. And worse, the eccentric millionaire can tell whether N has succeeded in forming

52. "The Toxin Puzzle," *Analysis* 43, no. 1 (January 1983): 33–36.

the desired intention by interpreting his brain states, and insists on paying N well before the toxin is to be drunk. Kavka thinks that it is at least very doubtful that in such a bizarre situation the unhappy N *can* rationally form the intention to drink the toxin. For as N thinks ahead to the time of the action, he can see that he will then have a serious reason not to drink the toxin and no reason whatsoever to drink it.[53] And foreseeing that he will be in such a state, he cannot with justification form the intention to drink it.

Kavka's doubt is very plausible. Indeed, I feel convinced that under the terms of the case N cannot rationally form the intention to drink the toxin. But if the case of punishment contains a viable justificatory structure, why isn't it also present here? Why couldn't N simply regard forming the intention to drink the toxin as an item of prior justification that will lend its justification to the later act? Our sense that drinking the toxin would not be rational seems to show that it is at least very doubtful that things can be conceived in this way; and this may suggest that there is something amiss in the very idea of actions being derivatively justified by reference to earlier actions or conditions that refer to them. But this suggestion rests on a false assimilation. The sphere of purely prudential rationality unconstrained by morality has features that rule out the special structure present in punishment. When prudential rationality is at issue, one is not, I think, able to separate the question of an action's justification from the question of the reasons one has for doing it. If one's reasons are good enough the action is prudentially justified; otherwise, it is not. In the toxin case the benefits that attach to forming the intention can provide no reason to do the intended action and thus cannot make it acceptable even in prospect.[54]

This difficulty need *not* arise, however, when the question of justifi-

53. N is not allowed to induce false belief in himself or to provide himself with independent moral motivation by, e.g., promising someone to drink the toxin.

54. David Gauthier in his recent paper "Deterrence, Maximization and Rationality," *Ethics* 94, no. 3 (April 1984), presents a view in which, so far as I can see, it would be rational for N to drink the toxin given the benefits that attach to forming the sincere intention to do so. For Gauthier advocates assessing the rationality of individual actions by first assessing the rationality of the largest temporal stretches of activity in which they occur (see p. 488). What he fails to make clear, at least to me, is how an agent who follows this policy is to think of his reasons. Is N at the later time to think that he has a good reason to drink the toxin despite the fact that no good will come of it? And if so, what does this good reason amount to? Or is N to think that this kind of choice can be rational in the absence of any reasons to make it? Neither option seems to me inviting.

cation is moral. For a moral justification need not be a function of one's reasons for acting. I am morally justified in reading your book because I have obtained your permission to do so; but my reason for reading the book is certainly not that I obtained this permission but that I hope thereby to amuse or instruct myself. The moral justification for an act of punishment does not have to lie in the punisher's reasons for punishing. Nor does his justification have to provide him with reasons to punish. Motives that do not in themselves morally justify an action can nevertheless constitute one's real reasons for doing it, and be perfectly acceptable in this role. In the case of punishment, such motives are not hard to find. Among them are those in which the two standard theories try to find its justification, righteous anger, and a desire to maintain the deterrent credibility of the penal institution.[55] Even more conspicuous, at least in complex practices like our own, are those mundane motives that arise from the fact that those who punish are expressly charged and employed to do so. Given that the punisher is in some sense aware of the justification provided by the right to make the earlier threats, he cannot be blamed for acting on reasons provided by any of these motives. And because we, as threateners, can foresee that we, as punishers, will have such reasons for punishing, it can be fully rational of us to form the collective conditional intention to punish.

Other objections could be made to this conception of the moral relation

55. The present theory can therefore accommodate some of the claims of retributivism as an account, not of the right to punish, but of a morally legitimate rationale for exercising part or all of that right. Nothing I have said so far implies that the natural desire to make wrongdoers suffer, given that one had the right to do so, is contrary to moral virtue. And its moral acceptability is suggested by the fact that benevolence does not seem to condemn us in taking some satisfaction in evils that wrongdoers suffer as, e.g., accidental results of their crimes. These natural attitudes must, of course, be held within certain bounds; otherwise they become cruel and vindictive. And perhaps the appropriate limit for *this* part of morality is some version of retributory equivalence. Note that the motive for punishing provided by our righteous anger, like that provided by our prudent desire to preserve the credibility of our penal institution, would not seem to generate the obligation to punish in any particular case. But that obligation might sometimes arise from other considerations. I see no reason, for example, why a penal code might not rightly mandate punishment for certain crimes or why authorities might not rightly promise the general public to punish in certain kinds of case. Moreover, I find myself strongly attracted to the idea that punishment of a crime can express the value society attaches to its victim and to his violated rights, and that not punishing or punishing too little may, in some cases, do the victim or his memory a moral injury. (For a discussion of other things that punishment might express see Joel Feinberg's "The Expressive Function of Punishment," *The Monist* 49, no. 3 [July 1965]: 397–408.)

between threatening and punishing. But here, as in earlier parts of the discussion, limitations of space have forced me to set aside some interesting problems and complications for other occasions. I have tried only to present a forceful sketch of an overall line of defense for what I consider a plausible but largely ignored theory of punishment. The heart of the theory is, as we have seen, the special justificatory structure described in this section. This structure may be, and I think is, present in moral and quasi-moral phenomena other than punishment. But in no other part of morality is its presence more plausible or, given that it is valid, its recognition of greater practical importance. I say this because I not only believe, as my objections to the standard theories have indicated, that punishment has been misconceived philosophically, but also that it has suffered from these misconceptions in practice. Our major mistake, I have argued, is to have focused too much on the punishing and too little on the creation of the threat of it. My hope is that with the correction of this faulty focus, we may be able to see that punishment requires of us neither an act of faith in the justice of retribution nor any neglect of rights for the sake of effects.