Death and Retribution

CLAIRE FINKELSTEIN

I Introduction

It is often supposed that a theory of punishment predicated on desert lends support to the death penalty. What leads to this assumption is a prior thought about the appropriate punishment for murder. If we are to punish murderers as they deserve, we will inflict on them what they inflicted on their victims, namely death. This association between a desert-based theory of punishment, known as retributivism, and the death penalty appears not only in academic writings on the subject, but in popular views of punishment as well. Public rhetoric in support of the death penalty, for example, is nearly always retributivist. Politicians urging its use in a particular case will more readily speak of justice and desert than of future dangerousness or setting an example for others. They evidently think the retributivist argument for death more appealing than the utilitarian arguments that might be made in its favor.

In my view, however, the faith that death penalty proponents place in the retributivist theory of punishment is misplaced. In this essay I argue that retributivism fails to justify the use of death as punishment, and, moreover, that a desert-based theory of punishment is particularly ill-suited to such a task. I shall not argue against retributivism as a theory of punishment per se. Although there may be reasons of a more general nature to reject retributivism, I will not attempt to make the more general case here. My more limited suggestion is that even if retributivism succeeds in justifying the practice of punishment overall, it cannot provide a compelling reason for including the penalty of death in that practice.

The present essay constitutes one piece of a more general argument against the death penalty. If I am correct that capital punishment cannot be justified on retributivist grounds, the death penalty proponent would need to find a different basis on which to argue his case. The most promising alternative line of argument appears to be general deterrence: The death penalty is justified against this murderer, he will have to argue, in order to deter other people who might murder in the future. But as I argue elsewhere, even if we grant the questionable empirical assumption that the death penalty really does deter, it is extremely difficult to justify the death penalty on the basis of deterrence alone. If I am correct that retributivism does not lend support to the death penalty, the death penalty proponent will find it substantially more difficult to argue the merits of that form of punishment, for he is unlikely to find support for his position elsewhere in political philosophy.

The implications of this larger claim for the death penalty are worth spelling out. If the death penalty cannot be affirmatively justified, it cannot be permissibly imposed. Why? More specifically, why not simply be agnostic about its moral permissibility, and allow practical considerations to determine whether we should employ it? The answer has to do with a fundamental assumption about the nature of punishment, namely, that punishment is a harm or evil to the person on whom it is inflicted. Although this assumption is a standard one in the literature on punishment, its implications have not been adequately noticed by legal philosophers and criminal law scholars. If punishment, including the death penalty, is an evil that stands in need of justification, then the burden is on proponents of the death penalty to justify its use, otherwise the death penalty must be assumed to be impermissible. Thus this relatively uncontroversial assumption about the nature of punishment has powerful implications for a debate about the moral permissibility of the death penalty: the death penalty may not be permissibly imposed unless it can be affirmatively justified. It also follows that the death penalty opponent need not show that the death penalty is morally unacceptable to make his case. All he need do is exploit the presumption that this is so, and then consider the adequacy of arguments to defeat this presumption. If the argument against retributivism I offer here is correct, then, the death penalty proponent will face a significant, and possibly insurmountable hurdle.

Claire Finkelstein is Professor of Law and Philosophy at the University of Pennsylvania.
II The Basic Retributivist Argument for the Death Penalty

Retributivism is the theory of punishment that asserts that punishment is justified because, and only to the extent that, the criminal deserves to be punished in virtue of the wrongfulness of his act. Traditionally, the core of the retributivist's argument for any specific penalty is the doctrine of lex talionis, the idea that a person deserves to experience the suffering or moral evil he has inflicted on his victim. Taken literally, lex talionis is an absurd doctrine—no one thinks we should rape rapists, assault assailants, or burgle the homes of burglars. This difficulty making sense of lex talionis has accordingly led some retributivists to suggest that retributivism is most compelling as a general justification for the institution as a whole, without thinking of it as containing a further theory of the measure of punishment.4 But in the absence of its accompanying doctrine of lex talionis, or any other way of giving content to the notion of desert, the retributivist will remain unable to justify any specific penalty, including the death penalty. Given that retributivism is absurd if accompanied by a literal interpretation of lex talionis, and vacuous (for our purposes) if articulated without lex talionis, the retributivist must attempt to cast his defense of the death penalty in terms of a more approximate system for matching crimes with punishments, one that does not insist that the punishment exactly fit the crime.

Most retributivists who argue in favor of the death penalty attempt to do just this. That is, they abandon lex talionis in favor of a similar idea, namely that a criminal deserves to suffer some approximate match for what he inflicted on his victim.5 But this approach turns out to be more problematic than one might have thought. Begin by considering how approximate the doctrine must be to work. It is not only that presently we are unwilling to inflict some of the more extreme harms, like rape and torture, that criminals sometimes inflict on their victims. The prohibited list also includes more modest harms, such as forcing a member of a fraternity to imbibe too much alcohol or requiring a rogue cop to remove his clothes and walk half a mile in winter along a public road, both harms that such perpetrators have inflicted on their victims. Indeed, once one begins to consider all the deviant forms of behavior our criminal codes outlaw, it is clear that the vast majority of criminal acts are not ones we feel entitled to impose by way of punishment. There are really only several criminal acts that we regard as yielding acceptable forms of punishment: false imprisonment, theft, and in some states murder. The retributivist who wishes to match crimes with punishments must come up with a theory that would limit the deserved penalty to the three forms of criminal conduct listed above.

There are two possible strategies available to the retributivist in order to accomplish this end. The first distributes punishments proportionately, so that the worst crimes are matched with the worst penalties, and so on down the line.4 This method dictates only relative levels of desert, rather than requiring any particular objective measure of what criminal acts deserve what treatment. We might call this version of retributivism the “proportionate penalty” theory.7 One can think of this strategy as an alternative to lex talionis or as an interpretation of it; the label is unimportant.8 For the sake of clarity, let us treat lex talionis as the theory that calls for a strict equivalence between crime and punishment. Proportionate penalty will thus be a modification of the basic lex talionis doctrine. Whatever its other merits, this latter approach will not help the retributivist to argue for the death penalty, or, for that matter, for any other particular punishment. For the method does not provide an argument that we ought to include a given penalty on the list of acceptable penalties. It merely insists on taking available punishments—that is, punishments we are already willing to inflict—and imposing them on perpetrators in order of severity according to the seriousness of the criminal acts performed. Thus, while the strategy is attractive in avoiding some of the difficulties that proponents of lex talionis face, it will not help the retributivist to argue for the death penalty.

The second, and more promising strategy is to attempt to establish a moral equivalence between crimes and permissible punishments. This strategy asserts that the perpetrator should suffer an amount equivalent to the harm or moral evil inflicted on the victim, but the kind of harm or moral evil involved need not match. That is, instead of either assigning the same harm or evil as punishment that the offender inflicted on his victim, or fixing penalties proportionately by making sure that the right intervals obtain between levels of punishments, we can match crimes with punishments on an absolute scale, but establish only a rough moral equivalence between the two. We would seek to inflict on the perpetrator by way of punishment the nearest morally permissible form of punishment to the act the perpetrator committed. Let us call this version of retributivism the “moral equivalence” theory of justified punishment.
Arguably, Kant suggested a moral equivalence approach to punishment on at least one occasion:

[A] monetary fine on account of a verbal injury, for example, bears no relation to the actual offence, for anyone who has plenty of money could allow himself such an offence whenever he pleased. But the injured honour of one individual might well be closely matched by the wounded pride of the other, as would happen if the latter were compelled by judgment and right not only to apologize publicly, but also, let us say, to kiss the hand of the former, even though he were of lower station.

The moral equivalence theory has no way of identifying which penalties are morally permissible and which are not.

That is, whatever other lessons we should draw from Kant's view on punishment, we should understand him as believing that a perpetrator can be treated in a way that is morally commensurate with the harm and suffering he inflicted on the victim, without having to inflict that very same punishment on him. And while Kant does not articulate the theory in this way, the basic strategy of such views is to try to distinguish what a person deserves, in some absolute sense, from what it is permissible for society to inflict on him by way of punishment. The moral equivalence theory thus maintains that while a criminal who locks his victim in the trunk of a car before killing her may "deserve" to be locked in a trunk himself before being executed, it is not permissible for us to inflict such a punishment on him. The moral equivalence theory suggests that we eliminate forms of treatment that are impermissible from our roster of available punishments, and within that constraint, attempt to match the offender's criminal act as closely as possible with the punishment we inflict on him. Once again, treating lex talionis as establishing a strict equivalence between crime and punishment, the moral equivalence theory is an alternative to the traditional doctrine.

But if this is the view we must take of the moral equivalence theory, that theory is woefully incomplete. For the theory by itself has no way of identifying which penalties are morally permissible and which are not. How do we know, for example, that locking a perpetrator in the trunk of a car and then killing him is impermissible under the theory, but that simply executing him is not? Without an account of which penalties are permissible, and why, we may as well argue that putting an offender to death is impermissible, but locking him up in prison for life is not, or even that lifetime incarceration is impermissible but that a twenty-year sentence is not. So the moral equivalence theory would need to be supplemented by another moral theory, one that would tell us which penalties are morally permissible and which are not. The theory of permisssibility then becomes a side constraint on the penalties it is permissible to inflict. But since the retributivist theory was supposed itself to answer the question of which punishments are morally acceptable and which are not, this would be a serious defect of the moral equivalence view.

Let us now suppose the moral equivalence theorist manages to supplement his view with an additional account that will tell us when a penalty is too harsh to be permissible imposed, and let us suppose that we accept the theory in that form. Still, it is not clear that the moral equivalence theory can be used to defend the death penalty. There are at least two problems with that view. First, even in this modified form, there clearly are some penalties we think of as morally unacceptable but which are less severe than death. And if we wish to rule out those penalties, we will be compelled to rule out death as well. Consider torture. It is difficult to see torture as off-limits on the grounds that it is unacceptably severe, because it is actually most plausible to think of torture as less severe than death. The moral equivalence theorist's own method makes it clear why this is so: If penalties are to be the equivalent of crimes, then we should rank penalties the way we rank crimes. But we think of murder as a more heinous crime than any non-lethal assault. So torture should be a less severe penalty than death. But if torture is an unacceptable penalty, it should follow that death is unacceptable as well, given that death is a more severe penalty than torture. Let us call this argument on behalf of the abolitionist the "severity response." It is his response to the moral equivalence argument, suitably recast in terms of a set of side constraints to eliminate certain very severe penalties from the pool of acceptable penalties.

Admittedly there are some problems with the severity response. First, the death penalty proponent might dispute the claim that death is more severe than torture. Torture is more severe, he might argue, because it is more uncivilized and more brutal than death. That torture is widely regarded by everyone, on all sides of the argument, as unacceptable but death is not seems to bear out the death penalty proponent's intuition on this point. But
what about the fact that most perpetrators themselves would choose torture instead of death? The death penalty proponent might say that the criminal’s own preferences cannot be the measure of the severity of a punishment. It is perfectly possible, for example, that a given criminal would prefer to spend a night in jail than to pay a small fine, because he is very attached to money and does not particularly mind confinement. But this would not show that the night in jail is a less severe penalty than the fine. There are also those who would prefer death to life imprisonment without parole. Yet surely these preferences do not imply that life imprisonment is a more severe penalty, even for those defendants.

Against this (and on behalf of the abolitionist), it might be noted that we can speak of the pain of having a tooth pulled and compare it favorably to the pain of having a severe burn, without its being infallibly true that everyone would choose the former over the latter. We are entitled to think of incarceration as “painful” because we assess that treatment in general terms. It need not be the case that everyone who has ever been “punished” by way of incarceration found that treatment painful.

There are many penalties we would readily classify as less severe than torture and death that we do not hesitate to say are impermissible to impose.

Finally, the death penalty proponent might object to the severity argument by pointing out that severity does not correlate terribly well with impermissibility. There are many penalties we would readily classify as less severe than torture and death that we do not hesitate to say are impermissible to impose. There is, for example, great resistance to the recent turn toward shame sanctions, such as forcing a convicted sex offender to display a sign outside of his dwelling revealing his status or forcing him to affix sex offender license plates to his car. It is also questionable whether it is permissible to sterilize repeat sex offenders, or to force female adolescent offenders to subject themselves to mandatory birth control measures like depo-provera. If the abolitionist’s severity argument were correct, however, we would have to take the lowest unacceptable penalty on the list of penalties and say that any penalty more severe than it would be morally unacceptable. This strategy would quickly rule out most sentences currently inflicted for felonies, since many objectionable shame sanctions are less severe than most terms of imprisonment.

On behalf of the abolitionist once again, and against this counter argument, we might say that perhaps we are wrong to reject shame sanctions and other minor interferences with liberty. Indeed, the above argument suggests a compelling reason to allow them, namely, that imposing them may enable us to avoid inflicting more severe penalties that involve significantly greater loss of liberty. If, for example, we have the choice between a shame sanction—like a sign or a license plate—and a period of incarceration, and if both penalties are equally effective from the standpoint of deterrence, we arguably have an obligation to inflict the shame sanction, since it is the less invasive penalty. On balance, then, the argument we offered above on the abolitionist’s behalf seems to me a good one—that if we reject torture because it is too severe, we should reject death as a penalty because it is more severe. This argument, however, seems to require that we revise our intuitions about a number of lesser penalties.

A second problem with the retributivist’s reliance on the (modified version of the) moral equivalence theory is that the argument he makes to defend the death penalty has also been used by death penalty opponents. Recall that the moral equivalence argument says that an offender deserves to suffer whatever he has inflicted on his victim, but that some penalties are simply impermissible to inflict, requiring us to find their moral equivalent. One of the most common arguments made against the retributivist death penalty proponent, however, is that there are moral side constraints on imposing the punishment of death, and that these operate over and above the constraints imposed by the notion of desert itself. Thus many death penalty opponents are happy to concede that a murderer in some objective sense “deserves” to die, but argue that death is not a permissible penalty for the State or for society to inflict.\(^a\) Death penalty opponents offer various reasons for this. Some maintain that society compromises its own civility by executing even the worst of our fellow human beings. Others argue that the State is usurping God’s role in deciding whether to put people to death, and that human beings are simply not sufficiently omniscient to pass life and death judgment on other human beings. Still others argue that there are certain penalties that are inconsistent with the requirements of human dignity, and that torture and death are among these.\(^b\)

What it is important to notice, however, is that the arguments for and the arguments against the death pen-
alty here have the same structure: both camps allow that someone who kills another person "deserves" himself to die, since offenders must receive the treatment they inflicted on their victims. And they both allow that we may not inflict certain penalties that offenders deserve, because some penalties are impossibly harsh. They simply disagree about whether death itself is such a penalty. Because neither side offers an account of severity or even particularly attempts to explain its stance on death, the debate between proponent and opponent seems arbitrary. It quickly reduces to the question of whether we think death an excessively harsh penalty—and that is not a terribly nuanced ground on which to settle the matter.

The upshot of this second point may appear to be inconclusive. The two sides of the debate reach different conclusions based on the same fundamental premises, and there seems to be no basis for picking one side over the other. But we should not be too hasty to declare the argument a draw. For the fact that we have no basis for choosing between the two sides should be understood as a benefit to the death penalty proponent. This follows straightforwardly from the assumption I articulated at the outset, namely, that punishment is a harm or evil that stands in need of justification. The result is that the death penalty proponent bears the burden of proof: The death penalty cannot be legitimately imposed unless we have some affirmative argument by which to justify it. Notice that the retributivist is particularly affected by this burden of proof claim, for by his lights killing a person is such an evil act that the killer incurs a tremendous moral debt, repayable only with the murderer’s own life. It would seem to follow that an executioner, or society more generally, who takes a person’s life must incur this same moral debt, unless his act is morally justified. In the absence of such a justification, the executioner, and society as his accomplice, is no better than a murderer.

What I am suggesting, then, is that the above arguments at least show that the retributivist has not met his burden of proof with the moral equivalence argument. Since I cannot meet that burden for him, I can only issue an invitation to the retributivist to make his case in greater detail. As I said at the outset, I suspect that retributivism as a theory of punishment is on strongest ground when it seeks to justify the institution of punishment as a whole, without attempting to justify particular punishments under that institution. But as such, it provides little assistance to the death penalty proponent.

III Conventional Retributivism

There is another version of the retributivist’s argument we must at least briefly consider, according to which principles of desert should be understood as grounded in contractarian agreement. This view enjoys a substantial advantage over ordinary retributivism. For the version of retributivism that is grounded in consent stands in an entirely different position relative to that presumption against punishment with which we began. Here the presumption, if anything, works the other way around: We seem entitled to assume, at least as an initial matter, that a treatment to which a person has consented is one it is morally acceptable to inflict on him. If it can indeed be shown that there is a consensual basis for treating the worst criminals as deserving of death, it will be much less difficult to justify inflicting it on them.

According to the view we are now considering, the reason the person who commits a particularly vicious murder deserves to die is that he has agreed to the norms that dictate this treatment. I suggest we call this "conventional retributivism," since this account makes the notion of desert, which lies at the heart of the retributivist approach, dependent on the prior agreement of those who will be affected by it. There are two versions of this kind of approach. According to the first version, we can actually think of the criminal as agreeing to his own punishment, since the criminal act itself constitutes the criminal's acceptance of the punishment that is later imposed on him. The criminal has intentionally done something that he knows is likely to result in a certain treatment at the hands of state officials. This is what I shall call the "voluntarist" theory of punishment. The central thought of the voluntarist account is that the criminal gives his tacit consent to be punished by performing the criminal act in the first place.

Carlos Niño, an expositor of this view, argues that assuming certain other necessary conditions are met (such as the fact that the punishment to be imposed is a necessary and effective means of protecting the community against greater harm), the performance of a voluntary act on the criminal’s part, knowing that he will be subject to punishment as a consequence, "provides a prima facie moral justification for exercising the correlative legal
power of punishing him."12 For Niño, then, the necessary consent is tacit, and it lies in the criminal’s voluntarily exposing himself to a certain penal sanction by performing an act he knows will, or may, result in his having that sanction imposed on him.

There are some problems with Niño’s account, however, brief consideration of which will prove instructive. First, it is not at all clear that a person who consents to run a risk of a certain consequence’s occurring can be said to consent to the occurrence of that consequence. So it is not at all clear that the fact that a person voluntarily commits a crime can be said to consent to the punishment he believes is likely to follow. It is not just that we cannot say that the consent a person issues to an act he completely controls does not carry over to its less than certain consequences. The problem runs deeper than that. It is not at all clear that the consent that attaches to the voluntarily performed act should be thought to transfer to a consequence of that act the agent foresees but does not intend. Arguably consent travels along intentional lines: the reason the voluntariness of the act implies consent is that the act (under that description) is intended. It would then seem right to think of any intended consequences of an intended act as consented to, but to think of other consequences as merely foreseen, without thinking of the agent as normatively committed to them.

A second problem with Niño’s account is presented by Larry Alexander.13 Alexander suggests that the voluntarist argument justifies far more than we are comfortable allowing by way of legitimate punishment. This is because justifications for punishment based on consent lack a principle of proportionality that would limit the level of punishment that could be imposed for a particular crime. Thus it would be perfectly acceptable to assign the death penalty for a minor traffic offense, Alexander says, as long as the offender was aware of the risk of receiving that penalty when he voluntarily broke the law. Although someone might be prepared to embrace this consequence of a consensual account, it seems a deeply objectionable feature of this theory, because it would put the consensual account out of sync with our prevailing practices of punishment. This is a problem that should worry the retributivist, for what it suggests is that the ordinary notions of desert that the retributivist thinks he is capturing when he argues that serious murderers should be put to death cannot in fact be grounded in contractarian considerations. Put otherwise, since retributivism trades particularly heavily on intuitions about which penalties “fit” with which acts, it does not appear to be open to the retributivist to dispense with proportionality concerns.

There is, however, a second version of conventional retributivism. On this version, instead of thinking of consent as operating act-by-act, and therefore looking for a way of establishing the criminal’s consent to the actual punishment he suffers, we should understand the entity to which the criminal consents as a general institution, or set of principles, which in turn provide a justification for a particular treatment of a given offender. The consent, that is, operates at the level of what Rawls calls the “basic structure,” instead of at the level of the institution’s particular response to the performance of a given prohibited act. As Jeffrie Murphy articulates this view:

On this theory, a man may be said to rationally will X if, and only if, X is called for by a rule that the man would necessarily have adopted in the original position of choice—i.e. in a position of coming together with others to pick rules for the regulation of their mutual affairs. This avoids arbitrariness because, according to Kant and Rawls at any rate, the question of whether such a rule would be picked in such a position is objectively determinable given certain (in their view) non-controversial assumptions about human nature and rational calculation. Thus I can be said to will my own punishment if, in an antecedent position of choice, I and my fellows would have chosen institutions of punishment as the most rational means of dealing with those who might break the other generally beneficial social rules that had been adopted.14

Justifications for punishment based on consent lack a principle of proportionality that would limit the level of punishment that could be imposed for a particular crime.

Let us call this second version of conventional retributivism the “consensual” theory of punishment. The question we must now ask is whether a consensual account of punishment would favor the death penalty. This is the question, to paraphrase Murphy, of whether individuals in an antecedent position of choice would have chosen the death penalty as part of the most rational means of dealing with those who commit the worst violations of the beneficial social rules that have been adopted. The argument that they would goes as follows. Each person enters into society because he fears for his bodily security and longevity. The security and life expectancy of each person is increased if those who violate society’s
primary norms are punished, for this will deter other potential criminals from violating those norms as well. Putative members of society would choose to include the death penalty in the schedule of available penalties for the worst crimes since, assuming the death penalty deters, they would thereby increase the expected benefits of a system of punishment. Let us call the effects on the bodily security and chances for longevity each person expects from a social policy or legal rule his "expected security." The argument for the death penalty, then, is that it increases each person's expected security.

Against the above argument, however, one should note that every member of society must also take into account the possibility that he will be subject to the death penalty, either because he chooses to commit a crime for which the death penalty is authorized, or because he is wrongfully thought to have committed such a crime. The death penalty thus also decreases each person's expected security. Thus the increased security provided by the deterrent benefits of having the death penalty must be balanced for each rational individual against the decreased personal security the death penalty also involves. Rational agents selecting rules for the governance of society would choose to have a death penalty if it turned out that the added deterrent benefits of having that penalty, over and above the deterrent benefits of life sentences, were greater than the decreased security to each person from the possibility of being subject to the death penalty themselves. Thus arguably it would be rational for contracting agents to include the death penalty in the rules governing punishment if its net expected security is positive. Now on the assumption that the death penalty does have significant deterrent effects, the net expected security it provides is probably positive. For a person presumably controls whether he will commit a crime and can choose to avoid committing one if he wishes to avoid death. And the chances of being erroneously subject to the death penalty seem considerably lower than the chances of being the victim of a crime that would be deterred by the death penalty.

But what about the person who is erroneously put to death? For him, there is no net increase in expected security. For as it turns out, his bodily security will have been destroyed by the existence of the death penalty. He would probably have done considerably better if he had lived with the increased threat of being the victim of a violent crime, and avoided the death penalty (even if he had to spend the rest of his life in prison anyway). And arguably since each person knows ex ante that he may end up in this position ex post, he cannot regard the death penalty as rationally motivated. For arguably it would be rational for parties to select a legal regime only if each one knows ex ante that he will be better off under the regime than if he had never selected the regime in the first place. And thinking that there are circumstances in which this will fail to be the case is perfectly consistent with that legal regime having positive expected benefits at the outset.

Whether rational agents in a contractarian scheme would agree to the death penalty under the circumstances is a complicated matter. If the ex ante expected security of the death penalty is positive, on one view this would be sufficient to make it rational for putative members of society to agree to it, even if the death penalty could end up leaving any given individual worse off than if he had not agreed to it. On another view, however, for putative members of society to agree to the death penalty, each person must believe that no matter how things turn out, he will be better off under the death penalty regime than he would be in its absence. In the present context, however, I cannot do justice to the debate between these different forms of contractarianism. Let us therefore avoid the issue by making a very implausible assumption, namely, that the deterrent benefits from the death penalty are so great as to leave even the person put to death under its rule better off than he would be in a society that did not include the death penalty in its available punishments. On this hypothesis, the difference between the two forms of the consensual account would disappear, and we have reason to think that the consensual approach to punishment would endorse the death penalty.

---

On the assumption that the death penalty does have significant deterrent effects, the net expected security it provides is probably positive.

The consensual argument for the death penalty appears to have a number of advantages over the voluntarist conception. In particular, it improves upon the earlier, voluntarist version with respect to two difficulties we considered above. First, there is no problem with consent to unintended consequences, since the voluntary nature of the act is now the source of the relevant consent. The voluntariness of the criminal's act serves to establish responsibility, and the fact that the criminal is responsible for a forbidden act makes him liable to punishment because he has consented to a scheme of punishment that
associates punishment with responsibility for harm-infliction. Second, the proportionality problem also disappears. For the justification for punishment will depend on general rules and principles, principles by which individuals have agreed to be governed. And those principles can place limits on the forms of punishment to which an individual can be subjected, even when he is aware that he will be subjected to it as a result of his own voluntary act. They will tie the justification for punishment to a set of principles that connects the point or purpose of punishment with the seriousness of the infraction the criminal has committed. These principles will be substantive, not formal principles, and so will likely have some kind of proportionality condition built in.

But even if rational agents can regard their expected security as improved by the existence of the death penalty (and indeed even if the death penalty meets the stronger condition of leaving all members of society better off than they would be in its absence), it need not be the case that rational individuals would institute it. For there are deontological norms to which the retributivist is committed, and these arguably counsel against legitimizing an agreement with a person to put him to death. To see this, consider a different, but instructive example, the case of the "Kidney Society."

Suppose everyone were at some small risk of finding himself with two failing kidneys. In order to protect against the risk of needing a kidney and having access to none, a group of people enter into an agreement to supply one another with a kidney by lottery, should one of the members of the group end up in this situation. The terms of the agreement specify that if any member of the group finds himself needing a kidney to survive, a lottery would be held to determine who would supply that individual with the needed kidney. Once a person is chosen by lot to supply the kidney, he would have no choice but to yield, and a kidney could be removed, by force if need be, in order to supply the needed kidney to the person with dual kidney failure. In this case, we could say that there is no benefit to the person who must have his kidney removed as the upshot of the lottery. He is clearly better off with two kidneys than with one, so that from his perspective, the gamble has not turned out to be worth it. But there was a benefit to him in entering into the agreement with others in the first place. We can therefore hypothesize that if he thinks the danger of dual kidney failure sufficiently great, and the loss to himself of being the one to be chosen by lot to donate a kidney either sufficiently remote or sufficiently bearable, he may regard the terms of the Kidney Society's agreement beneficial on the whole, making it fair to enforce its terms when he resists its application to himself. Like the increased net security we hypothesized the death penalty might provide, the members of the Kidney Society enjoy an increase in the net expected health benefits they experience, even counting the "costs" of having to provide a kidney.

There are deontological norms to which the retributivist is committed, and these arguably counsel against legitimizing an agreement with a person to put him to death.

Yet there are reasons to suppose that forcibly removing an objector's kidney would not be justifiable on consensual grounds, for it is not at all clear that rational agents concerned to maximize their long-term well-being would embrace the kidney scheme, or that ultimately it would be rational for individuals to ensure against kidney loss in this way. For despite the fact that each person can regard himself as better off living in a society in which he has insurance against dual kidney failure, at the risk of having to provide one of two extraneous kidneys himself, rational agents might not prefer a social scheme that provides them with such benefits. Rational agents might not want to select the "meta-regime" in which it was permissible to attack a non-consenting donor in the way necessary to enforce the Kidney Society lottery, even if the donation were one that the donor had putatively agreed to accept by signing up for the Kidney Society. One has only to reflect on the extreme discomfort we would have enforcing the kidney donation agreement if the person who had drawn the short straw in the lottery objected. Would we be willing to enforce such an agreement, solely on the grounds that the unlucky lottery participant had benefited from the insurance such a scheme gave him up until now? It goes without saying that no court in this country would order specific performance for such an agreement. Even if organ donation agreements were not void as contrary to public policy, the most one could expect to win against a recalcitrant donor would be monetary damages for the failure to turn over his kidney.

Presumably what bothers us about enforcing a kidney lottery is that we think the person who draws the short straw in the lottery has rights to bodily integrity, rights that he did not, and could not, contract away in the initial
agreement to enter the lottery. While we might imagine contracting away some substitute for one's kidney—monetary compensation, for example—we cannot quite understand one's own bodily organs as themselves subject to voluntary renunciation. Indeed, a retributivist ought to have particularly strong objections to seeing rights to bodily integrity as the subject of a contract between rational agents, for it is usually because of the retributivist's view of the robustness of such rights on the victim's part that he is so confident of his judgment that the murderer deserves to die for violating them.

Notice that the conventional retributivist would not necessarily be in the same position with respect to agreements to forfeit natural liberty. For incarceration leaves the body protected and one's natural life extended. It allows for the continuation of plans and projects of at least a rudimentary sort, and does not ultimately make impossible one's continued defense and self-protection in the event that one's conviction has been wrongful. The strong system of defensive rights I have been claiming that members of a social contract would insist on would not be rendered impossible by agreeing to an incarceration lottery.

IV Conclusion

The retributivist defender of the death penalty faces the following difficulty. If he seeks to defend the death penalty by claiming that an offender deserves to suffer precisely the treatment he has inflicted on his victim, the theory will prove absurd and morally repugnant: We simply are not prepared to rape the rapist and assault the assailant. So he must try to find the "moral equivalent" of the offender's criminal act instead, ruling out immoral acts and restricting available penalties to morally acceptable forms of punishment. But which acts are morally acceptable? The retributivist must at this point look to some supplementary moral theory to answer this question; the moral equivalence theory on the face of it does not supply the answer. But in this case, the problem is that the moral equivalence theory is incomplete, and so it cannot tell us which penalties we should inflict for which acts. Moreover, even if we do accept the theory in this form, there are further difficulties for the death penalty proponent. Many penalties we currently consider morally impermissible, such as torture, are actually less severe than death. And that suggests that whatever method the moral equivalence theorist uses to eliminate torture could also be used to eliminate death. So unless the retributivist is prepared to allow that criminals can be tortured, it will be hard for him to employ a moral equivalence strategy to argue for the death penalty. And in the absence of a clear justification for using death as a punishment, the death penalty is morally impermissible, given that there is a presumption against the infliction of any painful, involuntary treatment.

There is, however, another possibility for justifying the death penalty along retributivist lines, and this is to combine the basic retributivist approach with a consensual theory of punishment. Criminals would be subject to an agreed upon roster of deserved penalties, allocated according to general moral principles they endorse. On this strategy, it would not be necessary to justify the death penalty per se, since a treatment to which an offender has himself consented is at least presumptively morally acceptable. The most compelling version of what I have been calling conventional retributivism sees the offender as having agreed to norms that govern the institution under which he is punished.

But it is very implausible that rational agents primarily concerned with protecting and prolonging their lives would consent to an institution of punishment that contained such a terrible penalty, even if it provided them with positive net expected security. For rational agents would be unlikely to agree to equip one another with the power to put members of their own society to death, despite the benefits in expected security we are now assuming they would garner. They would thereby alienate the strong rights to bodily integrity whose protection was their primary motivation for entering into civil society in the first place. In other words, we have reason to suppose that rational individuals concerned to protect their bodily security would proceed according to a system of rights to bodily integrity, rather than independently adopting measures that appear to increase security directly. It is, at any rate, the desire to protect this system of rights that arguably lies behind the retributivist's conception of desert for punishment in the first place.
I am grateful to Michael Davis, my commentator from the American Philosophical Association Eastern Division meeting, for comments on the version of this paper I presented there. Thanks also to Leo Katz, Stephen Morse, and Seana Shiffrin for comments on the earlier version, as well as to Amy Shellhammer for assistance with research.

1 Finkelstein, An A Priori Argument Against the Death Penalty (unpublished manuscript, on file with author). Alternatively, the death penalty proponent might attempt to show that there are contractarian reasons to favor the punishment of death, namely that individuals establishing institutions predicated on mutual advantage would choose to incorporate the death penalty into their system of punishment. But I think it can also be shown that contractarian reasoning would lead one away from the penalty of death, as we will see when we consider a contractarian version of retributivism below.

2 The only challenges to this assumption have come from those who maintain that punishment is justified only if it is a form of treatment, rather than a painful sanction. This is the "rehabilitative" ideal that was so popular in the 1960s and '70s. That view has mostly fallen out of favor in recent years.

3 It is generally accepted in philosophical discussions of punishment. See H.L.A. Hart, Punishment and Responsibility (1968). See also Peiman, For the Death Penalty, in L.P. Peiman & J. Reiman, The Death Penalty: For and Against 1, 5 (1998) (defining punishment as "an evil inflicted by a person in a position of authority upon another person who is judged to have violated a rule").

4 See M. Moore, Placing Blame 205-06 (1997).

5 See Murphy, Death and the Supreme Court, in J. Murphy, Retribution, Justice and Theory 232 (1979) (although Murphy is at least tentatively against the death penalty in practice, despite its apparent acceptability according to him on theoretical grounds).

6 There is an ambiguity here. Does "worst" crime mean the criminal act that imposed the most harm on the victim? Or does it mean the act that is worst from the moral point of view? The two may not be the same. As I cannot explore the issue on the present occasion, I shall assume that the precise way in which one crime is to be judged worse than another remains to be specified.

7 For a more detailed exposition of what I am calling "proportional penalty theory," see Reiman, Why the Death Penalty Should Be Abolished, in The Death Penalty, supra note 2, at 94 et seq. Reiman calls this "proportional retributivism."

8 Reiman treats lex talionis as the principle that punishment and crime should exactly match. Proportional retributivism thus diverges from lex talionis, on his account, but is consistent with a theoretical adherence to that principle because it maintains that although the offender deserves to suffer what he has inflicted on his victim, there may be constraints on the degree to which we may morally give him what he deserves, ibid.


11 Jeffrie Murphy, for example writes: "It is by no means clear that one can show respect for the dignity of a person as a person if one is willing to interrupt and end his most uniquely human capacities and projects. Thus... there is perhaps a case to be made that the punishment of death is degrading..." Murphy, Cruel and Unusual Punishments in Retribution, Justice and Theory, supra note 4, at 243 (1979).


14 Murphy, Marxism and Retribution, in A Reader on Punishment 56 (Duff & Garland, ed. 1994). On at least one interpretation, grounding a desert-based theory of punishment in a contractarian moral theory is also the best way to understand Kant's approach to punishment. For it would help to reconcile Kant's emphasis on returning punishment for crime with the social contractarianism of his approach to legal rules more generally.

15 The latter view is, in my opinion, the better interpretation of the contractarian approach.

16 Arguably, the view I am now considering bears only a tenuous connection to the theory of punishment known as retributivism. Indeed, elsewhere I treat the view as a straightforwardly contractarian account. See my An A Priori Argument Against the Death Penalty, supra note 1. Here, however, I am treating the notion of desert as institutional rather than as a natural concept.

17 This is subject to the complication I discuss above about precisely how the contractarian approach is to be understood.