

The paradox of punishment and the reduction of desert

People should get their just deserts. In particular they deserve to be punished for their crimes. Of course, the punishment should be neither more nor less severe than the crime merits—the punishment should fit the crime. The more serious the crime the more severe the punishment deserved. A special limiting case of this is that those who are not guilty of any crime, the innocent, should not be punished at all.

On the one hand, these are all platitudes about crime and punishment, platitudes which, other things being equal, it would be better if a theory of punishment could both capture and explain. They are captured by retributivism of course, but unadorned retributivism does not offer an explanation. On the other hand, any reasonably sensitive person surely finds the whole business of institutionalized punishment problematic. More is required, by way of justification, than the repetition of retributivist platitudes. Gross injustices have been and are being visited upon people under the black umbrella of retributivist platitudes. In recent years, for example, hundreds of thousands of people in the land of the free have been incarcerated for possessing five grams of crack cocaine.

A wide range of non-retributive justifications have been invoked (education, rehabilitation, restitution, detentive prevention, the expression and communication of our values—to name a few) but perhaps the most obvious and promising of these is deterrence.¹ Rational agents weigh the expected costs and benefits of breaking the law, and since punishment may involve a substantial cost to them, they will factor that in.² An apparently devastating objection to this consequentialist justification is Goldman's paradox of punishment.³ Briefly, to serve as a rational deterrent, either the punishments would have to be disproportionately severe, violating certain attractive constraints on desert, or else the system of detection would have to be harrowingly intrusive, making a misery of the lives of the innocent and the guilty alike. I explore a broadly consequentialist rejoinder, one which initially appears implausible, but which promises to reconcile desert with deterrence in a very direct way.

1 Three approaches to desert

A theory of punishment may treat the concept of desert in one of three different ways.

Firstly, a theory can accommodate the concept by treating it as basic and underived, a primitive concept, not requiring any more fundamental considerations to motivate it. This is typically the position called *retributivism*, but I will call it *primitive retributivism* for obvious reasons. While the primitive retributivists might employ various metaphors to motivate their account, those metaphors should not be taken to be giving a deeper explanation of desert.

Secondly, at another extreme, a theory might simply reject the applicability of the concept of just deserts altogether. A theory might hold that the entire framework of desert is incoherent, even wicked, and should be abandoned. I'll call this *eliminativism* about desert. According to the eliminativist, the concept of just deserts should go the way of the concepts like *purgatory*, *phlogiston* and *witches*. If the eliminativist is right, it is never true that anyone deserves to be punished. I'll leave it to the eliminativists to make whatever case they can for totally undeserved but nevertheless justified punishment.

There is, however, a third possibility besides primitive retributivism and eliminativism. A theory may ground desert in, or reduce it to, other more basic concepts. Call this position *reductionism* about desert. The reductionist thinks there are truths about desert, like the platitudes with which I began, but that those truths reduce to other kinds of truths. A reductionist does not repudiate retributivist principles (as the eliminativist does) but he does not regard them as basic and underived (as the primitive retributivist does). Rather, he holds that retributivist truths are determined by, or supervene upon, other more fundamental truths.

It is important to recognize that reduction of desert is not elimination of desert, and that retributivist principles can be accommodated in these two quite different ways. In recent years a number of "reconciliationist", "mixed", "compromise", or "hybrid" theories of punishment have been proposed. These are theories which employ the retributivist concept of desert along with consequentialist considerations like deterrence. It would be wrong to conflate *reductionist* with *hybrid*, however. Every reductionist theory employs both sets of concepts, just like a hybrid theory, but the reductionist denies that desert is a primitive concept.

2 Primitive retributivism

Primitive retributivism maintains that there are primitive, irreducible facts about desert, and that there are primitive irreducible facts about the fittingness of punishments to wrongs perpetrated. This fittingness clearly involves principles of proportionality. For example:

Relative proportionality

The more serious a violation the more severe the punishment deserved.

In addition to relative proportionality, the primitive retributivist typically endorses individual claims of absolute proportionality, or commensurateness, as well. For example:

Deportation to a grim foreign land (viz. Australia) is a disproportionately severe punishment for stealing a loaf of bread.

A year's sojourn on a Pacific island (and subsequently being awarded various military honors by the French Government) is a disproportionately light sentence for killing a man while blowing up Greenpeace's *Rainbow Warrior*.

Since the primitive retributivist's theory of desert has to yield definite judgements on particular cases, it is presumably committed to something like the following:

Commensurateness

There exists a just desert function which associates each violation with the level of punishment deserved. The desert function is a well-behaved function of seriousness of violations.

“An eye for an eye and a tooth for a tooth” expresses one reasonably clear desert function satisfying both proportionality strictures—we can call it the *equivalence* function. The suffering or harm rightly inflicted on the wrongdoer must be equivalent to the harm he wrongly

imposed on his victim. Yahweh apparently endorsed a version of the equivalence function in conversations with Moses. It has often been pointed out by Biblical apologists that the equivalence function is less harsh than functions previously applied. Before Yahweh's directive, it was considered reasonable to take a couple of eyes, or maybe a life, for an eye. Such functions may have satisfied relative proportionality, but Yahweh evidently found them wanting on the score of commensurateness. Even so, Christians (following Jesus's lead here), have criticized the equivalence function for being too severe, perhaps because it is too closely connected to vengeance.⁴ Consequently most retributivists think the equivalence function establishes an upper bound on punishment deserved. That is to say:

Equivalence constraint

For all wrongdoings, the punishment deserved can be no more no severe than harm done.

For certain sorts of wrongdoings it may be obvious what the harm done is, and hence what the equivalence constraint amounts to. For others, however, there appears to be no determinate answer to the question of what exactly the harm was, or to the question of what punishment exactly fits it. This was made painfully evident in the report of the the sentencing of Mark Manes. Manes illegally supplied guns to the Columbine students (Harris and Klebold) who went on a rampage killing thirteen people before they turned the guns on themselves. Here are excerpts from a report published in the *Denver Post*:

There were differing opinions on whether Manes ... will pay dearly enough with a six-year sentence for the sale... .

“No way was justice served” said Bruce Beck, stepfather to Lauren Townsend, who was killed in the school library.

“The defendant put a hole in our hearts that can never be filled,” said Betty Hooks aunt of slain Isaiah Schoels. “Every day brings a new set of tears...If we had our way the defendant would never be allowed on the streets again.”

Sam Riddle, a spokesman for the Schoels family ...said the sentence was just. .. “It

was enough of a sentence - that's my intellectual response. Was justice served? Yes."

Some family members of the victims arrived at their own formula for justice. One figured Manes should serve as many years as the age of the oldest victim killed by the gun he sold. Another figured he should serve at least one year for each of the thirteen slain.⁵

Granted that Manes deserves to suffer for his crime, it isn't easy to see exactly what punishment fits it, as the varying responses illustrate. Furthermore, primitive retributivists will find it difficult to *motivate* a particular desert function without appearing to abandon their primitivism in favor of reductionism.⁶

3 Hybrid theories and the paradox of punishment

A primitive retributivist accepts standard platitudes about desert without any further justification, while a reductionist takes something else, like deterrence or rehabilitation or compensation to be primitive, and desert derived. A hybrid theory endeavors to exploit and combine independently given considerations—say, of both desert and deterrence. It is these theories for which the paradox of punishment is particularly acute.

Suppose primitive retributivism: that there is a correct desert function, specifying the appropriate and fitting punishments for all wrongdoings. Let us take a rational person who for some reason or other has lost the motivation to comply with the laws, and is tempted to perform some wrong despite a justifiable law against it. That is presumably because the action carries with it a certain (expected) payoff for the would-be wrongdoer. Moreover, let us assume a *parity* condition: that what the wrongdoer would gain from the violation is equivalent to what his victims would lose from the violation (not totally implausible, for example, in certain cases of theft). Further suppose that, absent any punishment, he would in fact perform the violation. The penalty will deter him if the deterrent effect of the penalty outweighs the payoff. The deterrent effect of the penalty assumes a maximum when the probability of detection is 1 and is nonexistent when the probability of detection is 0. Further, it is clearly greater the greater the probability of detection. The principle of expected value tells

us that in fact it is a linear function of probability:

Deterrent value of penalty = probability of detection X disutility of penalty.

The deterrent will be effective just in case this outweighs the payoff. But, given the equivalence constraint it cannot do so. Even on the worse case scenario, where detection is certain, the disvalue of the penalty exactly matches the payoff. In such a case the rational person will simply be indifferent between violation and conformity. So the deterrent cannot be effective. Even if we relax the equivalence constraint, so that the punishment can be a little worse than the harm done by the violator, for the punishment to be an effective deterrent we will still have to have a nearly infallible system of detector of violations, and a legal system which almost always convicts the guilty. That means we would all, guilty and innocent alike, be the subject of constant surveillance. Furthermore, to secure the high rate of correct convictions the presumption of innocence would have to yield, carrying in its wake a higher probability of wrongful convictions.

Of course, the assumption of parity—that what the lawbreaker gains is roughly equivalent to what the victim loses—does not generally hold. Even in what might be considered the clearest candidate for parity—namely theft—parity may fail. Consider the embezzlement of pension funds—the total marginal utility of the amount stolen may be far greater for the relatively poor victims than for their rich embezzler. But this fact about the contingent relation between payoff for violator and harm done to victim doesn't improve the prospects of the hybrid theory. For punishment is supposed to be a fixed function of seriousness of the wrongdoing, itself a function of harm done. However, if payoff can be a contingent and variable matter, then, given a certain detection rate, the effectiveness of a given punishment for a certain violation will also be rather variable, undermining there effectiveness if deterrence.

Deterrence can be made more effective by raising the severity of punishments to levels which typically outweigh the payoffs. For example, given parity, to achieve an effective deterrent with a detection rate of 1/2, the punishment would have to be twice as bad for the violator as the harm his victim had suffered. But that is not an option for the hybrid theorist,

since he claims that the punishment which a certain violation merits is independent of contingent matters like the actual detection rate, and that it would be unjust to inflict a more severe punishment just because that is what is required for deterrence.

How then, can we accommodate both desert and deterrence? It would be strange if our justification for punishing did not involve deterrence at the ground level. It is hard to believe that a system of sanctions against violators has nothing to do with deterring them from violations. But for punishment to be just it must not be undeserved. The paradox appears to show that these two desiderata are in conflict.

4 A solution to the paradox: the reduction of desert

Any theory which takes desert to be a primitive notion faces two problems. The first is that of determining a specific desert function. The second, even granted the function, is that of meshing it with the demands of deterrence. The reductionist holds that the notion of desert can be reduced to non-desert based notions. A reductionist theory is also a hybrid theory, trafficking in both desert and deterrence, and ideally delivering principles of desert like the platitudes we began with. However the reductionist claims desert is not a primitive notion, that it can be reduced to other more basic notions. In this section I will sketch and motivate a preliminary version of reductionism which seems promising.⁷

We begin with the simple observation that a system of conventions is not only valuable but essential for regulating people's interactions where there is a mixture of coincidence and conflict of interests. A convention is a regularity in the behavior of members of a group in some recurrent situation-type, typically a situation-type requiring *coordination* of behavior if people are to follow paths of high value. Think of the necessity for a system of road rules. We can set up the convention that all should drive on the right, or that all should drive on the left. If a convention serves to coordinate collective behavior, directing it along tracks of higher rather than lower value, I will call it a *norm*. Given a system of norms we can define the deontic concepts of *right* and *wrong*, *permissible* and *obligatory*, all relative to the norm system. Elsewhere I have defended the view that the deontic concepts are naturally construed as relativized to systems of norms.⁸

Norms are sustained by people's beliefs and desires. People must believe that things will go better if the norms are generally observed, and will typically want to abide by them on the condition that the norms are generally observed. Some norms have a built-in "penalty" for violators (like driving on the right) but with others violators might stand to gain something, usually at the expense of the law-abiding, precisely because the norm exists and others are generally restraining themselves. (For example, someone who parks in a no-parking zone, designated as such because parked cars are dangerous there, can do so only because others are not violating the norm.) In such a case the system may require sanctions for it to be sustained. The sanctions will serve both to deter would-be violators, and to strengthen the confidence of the non-violators in the viability of the system and the utility of abiding by it themselves.

We can think of *desert* as belonging to the same family of concepts as *right* and *wrong*, *permissible* and *obligatory*—a deontic concept. If these other concepts contain a tacit reference to a system of norms, it is natural to assume that *desert* does too, that facts about desert depend ultimately on facts about value—the value of norms and their maintenance. Relative to an obtaining system of norms which forbids a certain action, it is wrong to perform the action. Those who violate the norms, without a legitimate excuse, *merit* or *deserve* the sanctions used to maintain the system. Desert, like obligatoriness and permissibility, reduces to non-desert-based facts about the value of norms and the necessity of maintaining them through a system of deterring sanctions.

Unlike the eliminativist, the reductionist fully embraces the concept of desert. He maintains that anyone who deliberately violates a norm in a system *ipso facto* deserves the penalty laid out in that system (absent excuses, mitigating circumstances etc.). Which punishment is deserved is determined by considerations solely to do with deterring people from norm violations. This is reductionism. The concept of desert gains a place in our deliberations and justifications, but it is not, as a hybrid theory would have it, an *independent* factor working alongside deterrence in legitimating punishment, and hence potentially in conflict with it. Rather, desert is ultimately captured and explained by other considerations, notably deterrence. Further the shape of the desert function is a contingent affair—it is settled by contingent facts about what deters would-be violators.⁹

We can see immediately that one hoary old complaint leveled against consequentialist justifications of punishment does not apply to reductionism. It is conceivable that, on occasions, making an example of someone known to the authorities to be totally innocent, or imposing a much harsher penalty than the one which has been publicized, might enhance deterrence. However, the reductionist will not be forced to say that that would be the right thing to do, or that the suffering imposed would be deserved. Right and wrong, permissible and forbidden, deserved and undeserved, are all concepts defined by the system of norms and associated sanctions for its maintenance. It is impermissible to violate the norms, permissible to punish those who do, impermissible to punish those who do not, or to impose more severe punishments than the system prescribes. Those who violate the norms without excuse *deserve* the full burden of the penalties specified for that violation, but no more. Those who do not violate the norms don't deserve to be punished at all. Whatever the consequential effects of making this particular innocent person suffer, such suffering cannot be justified in terms of this reductionist theory of desert.

5 Desert under equilibrium

Does it follow from this sketch of reductionism that whatever penalties are instituted to maintain a system of conventions are just, and that violators of the conventions *deserve* whatever they get coming to them?

As a thirteen year old attending a high school in New Zealand, I was punished (*viz* beaten) by a gangly, red-headed, fundamentalist mathematics teacher—"Bloodnut" Tomlinson. Bloodnut disliked the sound of a pencil falling on the wooden floor—he thought it ruined the learning experience for the boys—and he had been steadily escalating the punishment for this offense as the term wore on. I was the first hapless boy to drop his pencil on the floor the very day that the punishment had escalated to caning. Two strokes. (Not many, I suppose, but enough to make it one of the most unpleasantly memorable days of an otherwise fairly happy childhood.) Bloodnut had a system of conventions that he was trying to enforce by penalties. He maintained a near infallible detection rate (he had very acute hearing), and he was simply raising the penalty to the level at which it would actually be an effective deterrent. Did I ever

drop my pencil again in math—well, no. Did I deserve to be beaten? I don't think so.

We need to say more to avoid the Bloodnut objection. Not any old system of conventions and penalties for enforcing them generates *just* deserts. Both systems of conventions and their associated sanctions are open to criticism. For a start, the conventions really have to be what I have called *norms* before the associated penalty function becomes a desert function. That is to say, they are rules which coordinate behavior along high tracks of value. It is doubtful that Bloodnut's rules achieved that, and the same goes for many bad conventions. On the other hand, we really do require road rules and gun control to stop people killing each other on a massive scale. For penalties to count as deserts, the conventions that we are attempting to enforce thereby have to be good ones—briefly, what I have called *norms*.

While each system of norms and associated sanctions for violation carries a concept of putative desert in its wake, the system has to satisfy an extra condition for the desert function to be a *just* one. We get *just* deserts (or just *desert*) when a system of norms satisfies an equilibrium condition. Stated simply the condition is this: detection rates and penalties are set at a level sufficient to deter potential violators, securing the goods which are the purpose of the system of norms, without making things worse for violators and non-violators than would rival settings of those parameters.

We have to be a bit more explicit about who we are trying to deter. Most people generally respect the norms and the order they are designed to ensure, and are easily deterred even when tempted by self-interest to violate a norm. These folk scarcely need to be deterred at all, and focussing on them would incline us to set the penalties far too low, thereby encouraging rascals. But no deterrent is foolproof. There are always crazy people who cannot be reliably deterred in a predictable way, either because they lack powers of rational deliberation, or because they lack the normal faculties which translate rational deliberation into action. Rather, we are interested in deterring non-crazy people who lack sufficient regard for others: those who either do not recognise the claims of others in their pursuit of self-interest, or worse, place a positive value on seeing harm done to others. We could call these people *rational egotists*, provided we keep in mind that the label is meant to embrace those who

have other-related desires of a malicious variety. Other things being equal, the seriousness of penalties should be set at a level which will typically be sufficient to deter rational egotists.

Suppose, then, that we have a system of reasonably good norms, and that the penalties under the actual detection rates are effectively deterring rational egotists from violating the norms. But suppose, further, that the penalties *seem* harsh. We might well ask, are they *really* too harsh? Are they *undeserved*? What would constitute *undeserved harshness* on this proposal? Well, if the penalties were relaxed would the violation rate rise? If the answer is *no*, then the penalties are clearly too harsh. If we could obtain roughly the same deterrent effect with less severe penalties then the extra suffering serves no good purpose, it is unnecessary, and hence, according to reductionism, undeserved.

Suppose that the system is not working to deter certain norm violations. Should the penalties or the detection rate be increased? Possibly, but not necessarily. Suppose we do increase penalties and that has no discernible effect on violation rates. Then the extra suffering inflicted on norm-violators is again ineffective, hence unnecessary and thus undeserved. This seems to be what has happened in the war on drugs. Suppose one grants that the laws against drug use would, if adhered to, promote value. Still, it is pretty clear that incarceration has soared while the rate of use has apparently hardly budged. Note, however, the following possibility: a convention might be a good one in the sense that it would be valuable for everyone to abide by it, but successfully enforcing it through a system of sanctions would be far too costly. Far more suffering would result than would be avoided. This is especially so if the convention is one which would involve massive changes in people's ingrained habits.

The total repudiation of production, distribution and consumption of that most addictive and harmful of drugs—tobacco—springs to mind as an example of such. The evidence is that the benefits of tobacco consumption are massively outweighed by health and opportunity costs. We would almost certainly be massively better off if smoking were to go away altogether. But, given that a quarter of all adults are severely addicted to the stuff, we would have to contemplate massive penalties to enforce generalized abstinence from tobacco products. It is implausible that the suffering caused by such a system could be worth it overall. The penalties it instituted would thus be undeserved.

6 Punishment under parity

Among the primitive retributivist commitments are principles which proportion the severity of punishment to the seriousness of the wrongdoing. The seriousness of a wrongdoing is, at least in part, a function of harm done to victims. Can the reductionist capture and motivate a reasonable principle of proportionality?

Would this reductionist proposal tolerate punishments more severe than the seriousness of harms inflicted? It could well do so if, for example, the harm done is typically on a par with the payoff (parity) and it is the sort of violation that is in general difficult to detect. In that case the punishment will have to exceed the harm done if it is to be an effective deterrent. When some norm or other is crucial for securing valuable outcomes, detection of violations is uncertain, and the harm to victims of violation is roughly on a par with the benefit gained by the violator, then rather severe penalties are called for—penalties more harmful to the violator than the violation was to his victim. But that seems reasonable—if the penalty were less severe unscrupulous violators would, over the long run, get away with massive gains at enormous cost to those who promote value by abiding by the norms.

Is this a violation of commensurateness? Recall that commensurateness involves the seriousness of the wrongdoing. The seriousness of a wrong is partly determined by the harm done: other things being equal, the more harm done the more serious the violation. However, other things may not be equal. Suppose Jack wants to harm Jill for some reason, and to do so effectively he will have to violate a norm. He has two options. One by its very nature would be very easily detectable, but the other would be much harder to detect. Are these two norm-violations equally serious? My intuition is that the one which is harder to detect is the more serious of the two. And it is also the one which, on the reductionist theory, deserves the harsher penalty. Hence, if seriousness is partly a function of detectability, then this is not a violation of commensurateness, which demands that penalties directly track seriousness of the wrong, and only indirectly track the amount of harm done.

Parity follows from a more general principle about the connection between payoff and harm done: namely, that the payoff for a certain kind of norm violation is proportional to harm done. Call this weaker condition *payoff-proportionality*. Take a certain kind of norm-

violation for which payoff-proportionality holds. The penalty for such violations, if it is to serve as an effective deterrent, will have to vary with the amount of harm produced: the greater the harm, the more severe the penalty. Consequently, since seriousness is proportional to harm done (given the detection rate is held fixed), the punishment deserved will have to be proportional to seriousness of the violation. Thus relative proportionality falls out of the reductionist account, given equilibrium and payoff-proportionality.

Here is another interesting illustration of the way in which a retributivist platitude about desert is delivered by reductionism. Given equilibrium and parity (that is, the utility to the violator is roughly equal to the utility lost by his victim), the total disutility violators as a group are subjected to through punishment will not be less than the total disutility which they inflict on their victims.¹⁰ Further, as a group the violators will not end up better off than they would have been had they not embarked on violation.¹¹ While that seems fully in accord with retributivist platitudes, it follows, given the reductionist account of desert, from considerations pertaining to the value of norms and the necessity of deterring violators.

The reductionist vindication of this retributivist platitude requires parity. However, there are two ways in which parity can fail. The magnitude of the payoff to the violator might be greater than the magnitude of harm done to victims (briefly: payoff exceeds harm) or less (briefly: harm exceeds payoff). I'll consider these in turn.

8 Parity violation: payoff exceeds harm

Consider a *kind* of violation such that the magnitude of the harm to victims is smaller than the payoff gained by the violators: for example, a very clever computer scam in which a hacker diverts a very small amount of money (one cent, say) from other people's bank accounts into his own. Let's suppose that it is constructed in such a way that it is difficult to detect the perpetrator. Suppose, further, that no-one counts the loss of one cent as *any* kind of harm to them—it is below the threshold of financial harm. (This might seem problematic for reasons connected with sorites paradoxes, but let's put that aside, since such problems are quite general.) So no-one, we can suppose, is harmed on a single perpetration of the scheme. Still, we would want to make such schemes illegal, for obvious reasons, and violators who are

discovered and convicted would deserve to be punished. Given the difficulty of detection, if the punishment is to serve as an effective deterrent to rational egotists it will have to be very severe—much more severe than the cumulative disutility to the victims (which, by assumption, is zero).¹² Consequently someone who perpetrated this scam would deserve a punishment that is far more severe than the total suffering he inflicts on his victims. Despite the fact that the disutility of the punishment to the violator is more severe than the harm he visited on his victims, this seems right. Perpetration of the scam on a single occasion will harm no-one. But repeated violations will, of course, harm people. Obviously we need to deter such scams, especially if they are difficult to detect. So despite the lack of harm caused by any one violation, perpetrators need to be deterred and, if caught, deserve to be punished. So the deserved punishment will have to be a stiff one if it is to be serve as an effective deterrent.

Now imagine a scam for depriving two hundred million people of one cent apiece which is very easily detectable. Suppose only one attempt in a hundred thousand would escape detection, and this is in some way obvious to violators and victims alike. (I leave it as an exercise to the reader to construct a plausible story with this structure) A very mild punishment would thus be sufficient to deter the rational egotist . It is my desert-based intuition that the punishment that this second kind of scam merits is indeed smaller than that merited by the first. The detection rate clearly makes a difference to desert.

The primitive retributivist will have a hard time explaining why these two scams merit such different punishments without conceding that desert depends on contingent features involving detection and deterrence. So where parity fails in the first way (payoff exceeds harm done) we get results which, while they may not accord with primitive retributivism, seem nevertheless to be right about desert.

9 Parity violation: harm exceeds payoff

Now consider the case where parity fails because harm done vastly exceeds payoff. It seems that if punishment is tied to deterrence then, according to reductionism, violations that fall into this category merit relatively light punishments, since rational egotists will be sufficiently deterred by such.

For example, consider the cruel killing of Matthew Shepard, the gay student in Wyoming who was beaten, strung up on a remote country fence and left to die in freezing temperatures. The killers stole \$20 from Shepard's wallet, and defense lawyers sought to portray this vicious killing as an unintended byproduct of a robbery. To deter prudent egotists from stealing \$20 all you would need to do is set a fine in the ballpark of $\$20 \times 1/p$, where p is the probability of detection. Does reductionism entail that in this case the punishment deserved is a relatively trivial fine? If so then reductionism is radically out of kilter with core pretheoretic intuitions about desert. If anything at all is clear about desert, it is that the wilful torture and killing of someone like the inoffensive Shepard merits a very hefty punishment indeed, certainly life imprisonment, perhaps even the death penalty.

This raises a general worry about my account. The level of punishment for harmful violations should not be radically sensitive to what the violators were hoping to get out of it. In the Shepard case, for example, the punishment shouldn't depend on whether the murderers were hoping for \$20 or for \$2,000. But if the deserved punishment is tied to deterrence, then it seems desert will be radically sensitive to the perceived payoff to the violators.

Here's a possible response. Recall that norms constrain people to certain kinds of behavior in recurrent situation types. We institute punishments for certain *kinds* of acts. Torture and murder are going to be high on our list of actions we want to rule out, and to deter. Torturing and murdering people can, of course, be carried out for a wide range of different motives, with a wide range of expected payoffs to different violators in different circumstances. Clearly we do not want to have to finely individuate these various violations of the basic rule by *all* of the many different possible motives, setting up a different system of penalties for each. Rather we want to make it simple and obvious to the rational egotist that he would be ill-advised to indulge in torture and murder whatever motives he might entertain for doing so. Now murdering can bring with it a large payoff to certain would-be murderers, and we need to cover these potentially large payoffs when laying down the system of deterrents. So the penalty for torturing and murdering should be large even if the detection rate is high, and despite the fact that quite a lot of murdering is carried out for what appear to be rather trivial payoffs.

This response is unsatisfying. First, it just seems wrong that the deserved punishment in the Shepard case should depend on what might go on in *other* cases. I don't believe that Shepard's murderers deserve a severe punishment because *other* murderers *elsewhere* are seeking higher payoffs. Rather, the severity of the deserved punishment must be more closely tied to the harm perpetrated against Shepard himself. How can the reductionist capture this core retributivist platitude which lies behind the principles of proportionality?

Second, suppose there is a kind of act such that the payoff to perpetrators seems *always* to be trivial compared to the harm to their victims, not because of contingent features, but because of the very nature of the act. Rape suggests itself as an example. The magnitude of the devastation of the rape to a rape victim seems way out of proportion to any direct payoff to the rapist. Does it follow that rapists deserve only light punishments, simply in virtue of the fact that rational egotists would be deterred from indulging in rape by the prospect of such? That hardly seems in accord with retributivist platitudes about just desert.

Rape is rather widespread despite the fact that penalties are not light, and that the detection rate is not negligible. This suggests that rapists, provided they are not simply irrational or crazy, do derive a considerable payoff from *something* connected with rape. Maybe they place a high value on the associated sexual sensations, but there are surely other less risky ways of procuring those. Perhaps, as feminists have urged, rapists get a kick out of exercising violent power over their victims, demeaning them in the process. If this is right, what the rapist values is thus directly and perversely tied to the harm he inflicts on the victim through coerced sexuality. Since what the rapist values is violently demeaning his victim in a sexual way, the greater the sexual indignities perpetrated on the victim the greater the payoff to the rapist. So it isn't at all obvious that rape fails to satisfy payoff-proportionality. What the rapist (perversely) gets out of raping is directly proportional to the harm done to the victim. Given this, it is possible to derive a clear justification for tying the severity of the punishment for rape to its degree of nastiness. Since it is precisely the nastiness which the rapist values, and the nastier the rape the more valuable it will be to him, the severity of the the punishment will have to be tied to the severity of the rape precisely in order to deter.

10 Proportionality

This last point, whether or not it applies to the case of rape, can be generalized to yield a general principle of proportionality.

Let us call an egotist *self-regarding* if he has only self-regarding desires, and cares about the desire-satisfaction of others only in so far that bears on his own desire-satisfaction. A self-regarding rational egotist will be easily deterred from perpetrating violations with low payoff, irrespective of harm to the victims. Thus if we only had to deter self-regarding egotists, parity violations of the second sort would indeed constitute an objection to the theory.

Call an egotist *malicious* if part of what he wants is that certain others suffer at his hands. For the malicious rational egotist the payoff from rape, torture, murder and other such violations includes the harm which the victims suffer through the violation. The more serious the harm done to the victim, the bigger the kick for the malicious rational egotist. It is plausible that Shepard's murderers, for example, were malicious, rather than merely self-regarding. Apparently they wanted to make Shepard suffer and die for being gay.

Unfortunately we have to deter malicious egotists as well as the purely self-regarding. A malicious egotist will be tempted by certain violations solely in virtue of, and in proportion to, the harm they bring upon their victims. Thus to be an effective deterrent to malicious rational egotists, the severity of the punishments will have to be tied in part to the severity of harm.

What about commensurateness? This adds to relative proportionality the claim that the desert function is uniquely determined. If this is the claim that there is one admissible desert function which assigns punishments totally independently of a system of norms, of levels of detection, and of considerations of overall utility, then clearly the reductionist will have none of it. But is that claim independently plausible? I don't think so. Suppose, however, that these other parameters (norms, detection rates etc) are held fixed. Then it is quite possible that, in conjunction with these settings, the equilibrium condition will pin down fairly narrow and determinate ranges of disutility for punishments. Finding out what the appropriate levels are will, of course, be a difficult business, involving empirical investigation and perhaps quite a bit of trial and error. But that is much plausible than the primitive retributivist's implicit claim that determining the appropriate level of punishment should be a purely *a priori* matter.

11 What do the criminally insane deserve?

Not all violations of norms deserve the penalty specified. There can be grounds for waiving a penalty, or setting it at less than the maximum. For example, no-one thinks that the insane deserve to be punished for norm-violations. However, insanity has quite a bit in common with various other failings in perfect rationality, and many such failings we think do deserve punishment. It is a virtue of reductionism that it helps to clarify these cases in a way which eludes both primitive retributivism and eliminativism.

Suppose a system of punishments is in equilibrium so that it is effectively deterring rational egotists at an appropriate level. Won't irrational or stupid egotists be tempted to commit violations in droves? And aren't irrationality and stupidity too close to insanity for comfort? If we don't excuse mild forms of irrationality and stupidity why should severe forms of irrationality and stupidity be excused?

There could be a number of reasons why an otherwise rational egotist would violate a norm in the face of a deterrent which satisfies the equilibrium condition. Firstly, he may know the penalty and the detection rate, but the payoff for him on that occasion might outweigh that. Clearly we can't rule that out without making penalties so harsh that equilibrium is sacrificed. Secondly, an otherwise rational person can be too lazy or too preoccupied to carry out the requisite calculations. That's a kind of personal negligence. Thirdly, most people have a reservoir of personal recklessness which can overflow and drown out rational calculation. Even when the payoff is not worth the risk of the penalty, they may violate the norm anyway, perhaps taking the chance that they will get away with it. Finally, there are the insane who cannot carry out the calculation, or even if they could they wouldn't act appropriately in the light of that.

The first three kinds of violation seem worthy of punishment, but not the fourth. Why? Suppose we are setting up the system of sanctions to enforce our norms. We will have to decide what kinds of violations, if any, might be excused. Certainly we won't excuse those who make a rational calculation and consider the violation worth the risk of punishment. But consider insanity. The no-exception policy will not serve as a deterrent to the congenitally

insane, by assumption. So we would get no extra deterrent benefit from punishing insane violators, while the punishment will increase their suffering. Such a system would not be in equilibrium. The consequentialist would thus be motivated to make insanity a defense.

What about negligence and recklessness? These are lapses from full rationality. However not every person who acts negligently or recklessly is insane. We often say of ourselves that we *should* have controlled that lapse in rationality, and in so doing we regard ourselves as walking a line between rationality and temporary irrationality, sanity and temporary insanity, and that which side of the line we tread is at least partly up to us. We certainly want to discourage people from cultivating negligent and reckless habits, just as we try to deter children from cultivating a habit of throwing away self-control in wild tantrums. So in setting up our system of sanctions and legitimate excuses the reductionist will make a distinction between the congenitally insane, and those who lapse into such negligence and recklessness. And she will set the standard of proof of congenital insanity rather high so that people will not be encouraged to so lapse.

We want the system to maximize benefits, and there is simply no benefit, in enhanced deterrence or anything else, to making the insane subject to the same punitive sanctions as fully rational egotists, or in excusing those who fail to control their temporary irrational impulses. Given reductionism, it follows that the reckless and negligent do, while the congenitally insane do not, deserve to be punished.

12 Paradox revived?

Suppose we have a system in equilibrium, so that dropping the rate of detection while holding the penalties fixed, or reducing the penalty while holding detection rates fixed, would cause a non-negligible increase in violations. If this is right we could presumably maintain exactly the same deterrent effect by substantially reducing detection rates and compensating for that by substantially increasing the penalties. The new system would still be in equilibrium and would save us a lot on surveillance and enforcement. The enormously increased punishments would be thus also be deserved, despite being way out of kilter with the desert-based platitudes we have been trying to capture. Thus the paradox has an ugly reincarnation.

There are three reasons why this kind of adjustment might well not produce a system in equilibrium.

Firstly, when probabilities become small, and benefits or burdens large, even otherwise rational people are bad at carrying out the rational calculations, or acting on them. (Witness the popularity of lotteries.) If we aim penalties at perfectly rational egoists then a lot of egoists who fall short of full rationality will be tempted to commit violations because they can't handle expectations at low probabilities. So there will be more violations and hence more unnecessary suffering. Secondly, if the penalties are very high then in cases of a false conviction the suffering will, of course, be undeserved, and grossly so—much worse than the undeserved suffering of a false conviction when the penalties are lower. Most people regard an undeserved penalty as that much worse than the same penalty when deserved. So with the same rate of false convictions the higher penalties will cause more overall suffering than the lower penalty with the higher detection rate. Again the system will not be in equilibrium.

Thirdly, the fear of being wrongly accused and convicted, and hence the subject of such grossly undeserved suffering, even when it has a low probability, is itself a continuing and unnecessary burden on innocent, law abiding citizens.

13 Desert and reward

Some deserve to be punished for the wrong that they have done, but others deserve to be rewarded for the good that they have done. Some claim that Rosalind Franklin deserved the Nobel Prize for her role in the discovery of DNA. Others that Penzias and Wilson didn't deserve it for their almost serendipitous discovery of background radiation. Some claim that Mother Teresa deserves an eternity in paradise for her efforts on behalf of the poor. One cannot reduce just *half* a concept and expect to get away with it. Nothing in what I have said so far tackles the positive half of desert: rewards and honors. More work needs to be done on the better half of desert, and here I will only sketch considerations which suggest that the account can be generalized.

One can argue, in a parallel way, that a system of rewards and honors similarly presupposes a background system of norms which it is valuable to promote and sustain.

Rewards might serve as incentives in three related but distinct ways. One would be to encourage people to stick to the norms under difficult or trying circumstances. A second would be to encourage them to excel within both the letter and the spirit of the norms. A third would be to encourage them to surpass (without violating) the norms in their pursuit of value. All this seems a quite natural and plausible extension of the analysis of desert in punishment. It is not accurate, then, to say that desert reduces to deterrence. Desert, like obligatoriness, permissibility and the *supererogatory*, reduces to non-desert-based facts about the value of maintaining norms through a system of deterrents *and incentives*.

The Nobel Prize seems to fit rather nicely into this scheme nicely. There is a background system of conventions, many of them tacit, and not always well understood, about how the enterprise of science should be conducted. The Nobel Prize was instituted to encourage practitioners to strive for excellence in science within those conventions, and it certainly seems to have the desired effect. “The Big N” is a powerful motivator for serious scientists. Of course, they want it as much for the kudos as for the money, but these are closely related. If Franklin deserved the Nobel Prize for her contribution to genetics she thereby deserved her share of the prize money. In fact, she did not deserve the award in 1962, because she was dead by the time the prize for that discovery was awarded, and it is one of the rules that it cannot be awarded posthumously. (There is a genuine question about whether she would have deserved it had she remained alive.) But absent the rules and regulations of the institution of the Nobel Prize, it hardly seems true that a certain scientific discovery deserves an award of a determinate amount of money. The notion of just desert here is tied to the obtaining system of conventions, and the actual system of rewards designed to promote valuable work.

What about Mother Teresa? It *makes sense* to say that she deserves an eternity in heaven for her good works, but what would make that claim *true*? According to the reductionist, it would be true if God had set up the familiar system of punishments (hell) to discourage violation of his norms, together with a system of rewards (heaven) to encourage compliance with the norms, *provided* the resulting system were in equilibrium. If there is no such system then the claim is just false, albeit intelligible. In this case, we could charitably regard the claim as elliptical for a true counterfactual conditional: *if* the Christian God together with heaven and

hell had existed, then Mother Teresa would have deserved an eternity in heaven.

There is, of course, another possibly intractable problem engineering a system of infinite rewards and punishments consistent with equilibrium. An infinite disutility cannot be a justified deterrent of a violation with a finite payoff if the detection rate is greater than zero. Some finite amount suffering would always do the job, and so an eternity in hell would involve an infinite amount of unnecessary, *ipso facto* undeserved, suffering. An apologist for hell might point out that this is a case in which the actual detection rate (which is 100%) has to be distinguished from the subjective probability of detection (which may be *extremely* low, as it is by my lights). However this position faces another problem. For infinite suffering to constitute just deserts, the subjective probability of detection would have to be smaller than *any* real number without being identical to zero. In other words we would have to countenance infinitesimally small degrees of belief. A hell-creating God would have to make sure that those he condemns to hell walk this extremely fine line between adopting outright atheism, and conceding there is an infinitesimally small chance that God exists.

These problems generated by punishments of infinite disutility do not undermine this analysis of desert. Rather, in conjunction with the reduction of desert, they rationalize our pretheoretic discomfort with the whole machinery of final judgement and eternal damnation, which really does seem to violate any reasonable doctrine of commensurateness. Reductionism thus explains why even the rational faithful don't really put their money where their mouth is. In their hearts they know a system involving sanctions of infinite disutility cannot be in equilibrium.

There are, of course, notable differences between deserved punishments and deserved rewards. The role of detection seems minimal, even nonexistent, in the standard case of rewards. We do not generally think that the magnitude of the reward deserved should be sensitive to the probability of finding the people who are worthy of it. Why this asymmetry? The reason is, of course, that violators of norms are generally striving to avoid detection and the consequent punishment. Violators will typically exploit weaknesses in the system of detection. By contrast, those who are exceeding the norms in the production of value are almost never trying to hide their light under a bushel. They have no stake in remaining undetected. Scientists are not trying to hide their discoveries from the Nobel Committee.

Mother Teresa wasn't terribly secretive about her activities. And even if Christian do-gooders are sometimes enjoined not to flaunt their good actions before others that is because the One maintaining the system of detection and rewards is sure to know all about them anyway. Other *humans* play no role. However, suppose a certain sort of valuable contribution were by its very nature extremely difficult to detect, and suppose further that we would be justified in trying to encourage it by a system of incentives. Then it seems we would be justified in making the associated reward very large. And in such a case it does not seem at all odd to say that the designated reward would be, *ipso facto*, deserved. (A plausible example of such a kind of action eludes me, but that merely reinforces the point that the asymmetrical role of detection in reward and punishment is rooted in the different natures of the actions to which they are tied.)

14 Freely-floating desert

Sometimes we make judgements about just deserts which appear to float free of any system of conventions or norms. A struggling solo mother, holding down two jobs, fighting breast cancer with no health insurance, uncomplainingly pouring her love into the raising of her three children, suddenly wins Lotto and her life is transformed for the better. We say: "She really deserved that million dollars!" Likewise, if she misses the Lotto by one number, and dies from the breast cancer after a long and painful fight, leaving her children homeless, we might say: "She didn't deserve a fate like that, especially after all the good she's done." What can the reductionist say about these judgements of apparently freely-floating desert? In neither case is the outcome part of a system of rewards and punishments for sustaining actual norms.

The primary notion of desert is, according to reductionism, anchored to the notion of maintaining and promoting a system of norms for regulating our interactions. What a person deserves *in fact* will depend on which particular system of norms (if any) is *actually* being implemented. However, no particular system of norms is part of the *concept* of desert, and there will be various general claims about desert which are true provided only that some or other system of norms is in force. These are the platitudes about desert, like those we began with: "People should get their just deserts. The wicked (those who deliberately violate the norms) should be punished. The wonderful (those whose goodness exceed the requirements

of the norms) should be fully rewarded.” and so on. It is these these platitudes which inform fairly general, free-floating claims about desert.

What about more specific claims? In an ideal world there would be norms, both tacit and explicit, which people would want both to abide by and to excel within. Violations would be rare, and punishments swift and effective. Striving for excellence would be common, and when achieved, recognized and generously rewarded. I conjecture that when we make more specific free-floating claims about desert we are thinking of desert under an ideal system, a system we would all recognise as conducing to flourishing and excellence across the board. Good people, striving to promote and enhance value under difficult circumstances, would not be left to suffer and die needlessly, while the wicked flourish. And even if we are not thinking of an ideal system, we are thinking of a system which improves on the actual one in various desirable ways. As things stand, there are no, or few, institutions for rewarding the solo mother for her exemplary efforts in the face of hardship, or for preventing her from suffering and dying from a preventable condition. But intuitively we feel that this is a defect in our social arrangements. We can imagine better arrangements, those in which, quite generally, those who do good really do flourish, and those who deliberately and wilfully do evil suffer for it.

15 Why the reductionist can't have it all

Can the reduction of desert to deterrence capture *everything* a primitive retributivist would want to say about punishment? Hardly. Primitive retributivists will deny, for example, that desert is grounded in deterrence, or that desert can be a function of detectability, and so on. But obviously these high-level theoretical disagreements should not motivate us to abandon reductionism. A good reduction will capture as much as possible of what we could call the *observational* content of the reduced theory—in this case, what I have called the platitudes about desert. But even here the reducing theory may shine its light on areas where the reduced theory gets things wrong.¹³ I hope to have illustrated some ways in which the reduction of desert to non-desert based notions can capture what is attractive and important in the platitudes that underlie primitive retributivism. Where I part company with the primitive retributivist will, hopefully, turn out to be precisely the point at which primitive retributivism is wanting.

Footnotes

- 1 For interesting recent work on punishment, see Simmons 1995.
2. For simplicity, throughout I will assume that the probability of detection is just the actual detection rate, and that this is common knowledge.
- 3 See Goldman 1979.
4. For a recent vigorous attack on the widespread presumption against the morality and rationality of vengeance, see Barton 1999.
5. From “Families find little comfort in Manes' sentence” in *The Denver Post* November 13, 1999, section A , p. 21.
6. Michael Davis, a retributivist who has devoted more energy to this problem than most has suggested two different methods for determining the correct desert function. See Davis 1983, 1992 and 1993 However Scheid 1995 has shown that the two methods fail to converge on a single desert function. Further, Davis’s market method for determining unfair advantage is apparently circular, since it both presupposes, and is sensitive to, an already established system of punishments.
7. This is not to say that it is the only version of reductionism, nor that this version is wholly original. It shares certain features in common with the celebrated hybrid theories of Rawls 1955, and Hart 1968, as well as with Quinn’s recent justification of punishment in his 1985 (see f.n. 9).
8. See
9. Quinn 1985 can be construed as arguing for a variant of this reductionist theory, although he does it without invoking a system of norms and he does not talk about desert.

According to Quinn it is legitimate (rational, moral) to threaten a person with certain sanctions to deter them from wrongfully harming you. However, if the threat fails to deter, then it is legitimate to carry through with the threat. According to Quinn it cannot be legitimate to threaten a sanction if wronged, and illegitimate to carry out the threat when the wrong transpires. The threatened punishment would thus be something that the wrongdoer cannot justifiably claim is wrong. If it is not wrong it is permissible. We can add that since it would not be permissible if he did not deserve it, he is getting his just deserts. Deterrence, when it fails, delivers just deserts.

10 Let U be the utility of the violation to the violator, and let d be the detection rate ($0 < d < 1$). Then (by parity) we can set the disutility to the victim at $-U$. For the punishment to be an effective deterrent its (dis)utility D will have to satisfy $dD + U < 0$. That is $D < -U/d$. (If d is 0 then no punishment will suffice as a deterrent and hence none will fit the crime.) Suppose there are N violations, thus yielding a total disutility to victims of $-NU$. Then dN violators will be caught and punished yielding a total disutility through punishment of $(dN)D < (dN)(-U/d) = -NU$.

11. For as a group they gain NU through their violations while they lose more than $-NU$ through punishment.

12. We could, however, easily construct less extreme cases in which total disutility to victims was non-zero but still substantially less than the disutility of the punishment deserved.

13 Galileo's theory of terrestrial motion and Kepler's theory of planetary motion were reduced to Newton's generalized theory of motion and gravity. But Newton's theory showed up errors in both.

References

- Barton, C., *Getting Even* (La Salle: Open Court, 1999).
- Davis, M. “How to Make the Punishment fit the Crime” *Ethics* 93 (1983), pp. 726-52;
To Make the Punishment Fit the Crime (Boulder, CO: Westview, 1992);
“Criminal Desert and Unfair Advantage: What’s the Connection?” *Law and Philosophy* 12 (1993), pp 138-9;
- Goldman, A. H. “The paradox of punishment”, *Philosophy and Public Affairs* vol 9 (no 1) (1979), reprinted in Simmons 1995, pp. 30-46.
- Hart, H.L.A. “Prolegomenon to Principles of Punishment”, in *Punishment and Responsibility* (Oxford: Clarendon, 1968), pp. 1-13,
- Quinn, W. “The Right to Threaten and the Right to Punish”. *Philosophy and Public Affairs*, 14 (1985) reprinted in *Punishment* ed. Simmons *et al.* pp. 47-93.
- Rawls, J. “Two Concepts of Rules”, *Philosophical Review* 64 (1955) pp. 3-32.
- Scheid, Don “Davis, Unfair Advantage Theory, and Criminal Desert”, *Law and Philosophy* 14 (1995), pp. 375-409.
- Simmons, A. John (et al) eds. *Punishment* (Princeton University Press, Princeton N.J., 1995).