## Taking the Bad with the Good:

### Some Misplaced Worries about Pure Retribution\*

#### Gerald F. Gaus

# Philosophy Department and Murphy Institute of Political Economy Tulane University

I

In his essay on "Criminal Justice and Legal Reparations as an Alternative to Punishment," Geoff Sayre-McCord notes that in an earlier paper I argued that compensation for a violation of another's rights could *not*, to use an Aristotelian notion, "restore the moral balance." Compensation payments, I argued there, may in some cases undo the harm done to the victim insofar as they can restore the victim's overall interests to the their pre-violation level. But, I argued in that paper, even if a person is fully compensated in this way, it would fail to right the wrong. Even after the offender fully compensated the victim for the harm done, she is, as it were, in moral deficit. At the close of that essay I conjectured that punishment may complete the job begun by compensation — it may help restore the moral balance — by returning us to a just situation. Sayre-McCord, along with many others, rejects this. Indeed, he finds it "mysterious, to say the least" how punishment could serve this function.

In this essay, then, I want to do two things. First, I want develop the conjecture I made at the end of this earlier essay, that punishment can restore the moral balance. Secondly, I want to try to locate the reasons why so many philosophers, including Sayre-McCord, have resisted this

<sup>\*</sup> This paper has been presented to the Philosophy Departments of Tulane University and the University of Arizona and, originally, to the 1999 Sociedad Filosofica Ibero Americana (SOFIA) Conference on Legal and Political Philosophy, in Mazatlan, Mexico. I am most thankful to all the participants. I am especially grateful for discussions with Julia Annes, Tom Christiano, Eric Mack, Geoff Sayre-McCord, David Schmidtz and Michael Smith.

idea, and the broader conception of retribution with which it is associated. My aim here is not simply to present one more entry in what often seems an endless series of papers about the merits and demerits of retribution. Although I think this is a significant issue, in my mind the more important philosophical task is to locate the source of this seemingly intractable philosophical dispute. As I indicate in my title, I believe that these worries are typically *misplaced*. They are understood as worries about the notion of punishment, but they run much deeper. Ultimately, I think they lead back to deeper disputes about deservingness, practical rationality, principled action and the role of the state. I do not want to deny that they are worries, but all too often they have been perceived as specific worries about punishment, when, I shall argue, they are better understood as worries of a more general nature.

#### II

To focus the discussion, let us deal with a simple, widely held, and perhaps even more widely criticized, conception of justified punishment, which I shall call:

*Pure Retribution*: (1) Criminals deserve to be punished; (2) punishment is intentional harming, with the aim of harming.

Let me note two points about this characterization of pure retribution. First, this is not the only understanding of retribution: a defense of retribution need not depend on a theory about desert, but a formulation along these lines is by far the most popular one in the literature. As C.L. Ten notes, "Contemporary retributivists treat the notion of desert as central to the retributive theory, punishment being justified in terms of the desert of the offender." Although some critics — and a few proponents — of retribution seek to depict it simply as an institutionalized form of revenge, it seems fairest to understand it as a claim about deserts.

Second, I call this theory "Pure Retribution" as it does not understand punishment simply as the intentional infliction of harm, but as an act that *aims* at harming the offender. Sayre-McCord defends a theory of reparations, according to which offenders loose their status as full members of the community, and so can be denied certain rights and privileges. Now although this denial is an intentional act that typically results in harms to the offender, the *aim* is not to inflict harm on the offender. Some of the offender's rights and privileges are withdrawn because, as it were, she has lost the status of full member, and this may harm her, but it is a foreseen, but not sought-after, result of our negative treatment. Sayre-McCord thus employs a version of the doctrine of double-effect. Although a side-effect of punishment may be harm to the offender, that is not its aim or *telos*: the harming is a foreseen, but not sought-after, effect of the action. In what I call Pure Retribution, harming the offender is the aim of punishment.

In addition to Pure Retribution, I want to defend:

*Moral Restoration*. In some cases of criminal offense, punishing the offender is a necessary condition of a return to "equality, by which the pointer of the scale is made to incline no more to one side than the other." There exist some cases in which punishment, along with compensation, is part of a set of necessary and sufficient conditions for this result.

This claim, which I am drawing from Kant, is essentially metaphorical; we need to explicate the metaphor, but something like it has been a crucial feature of disputes about retribution.

Advocates have consistently employed talk about restoring the moral balance or annulling the crime. And critics have consistently thought it is thoroughly muddled. Note, though, that my statement of the Moral Restoration avoids the easiest objection. It of often charged, for example by Ten, that the Moral Restoration claim confuses punishment with compensation or reparations.

My concern here is only with the moral situation after all compensation payments that are possible have been made. As I said, in my earlier paper I argued that even after all possible compensation has been paid, and so the harm has been undone, the wrong remains. So I am concerned with whether punishment can, in some sense, undo the residue wrong, not whether it can undo the harm done.

Together, Pure Retribution and Moral Restoration account for the crux of what many people wish to affirm when they uphold retributivism. However, I think one additional claim is necessary. Recall Hobbes's definition of punishment in *Leviathan* according to which a punishment is "an evil inflicted by public authority, on him that hath done, or omitted that which is judged by the same authority to be a transgression of the law." In one way, of course, this clearly isn't right. We don't want to say that, as a matter conceptual necessity, punishment must be meted out by public authorities; some parents do punish. But almost all the discussions of punishment and retribution suppose that meting out punishment is a possible, indeed basic, task of public authorities. If somehow it is impossible for public authorities to punish, even a defense of the Pure Retribution and Moral Restoration claims would have little significance for debates about law and punishment. I thus wish to defend a third claim:

Political Possibility: It is within the competency of public authorities to punish offenders by giving them their just deserts.

#### Ш

All three of the claims I shall defend often have been the subject of criticisms — or, at least, the source of worries. First, and most familiar to all, Pure Retribution is often said to be inhumane, barbaric or simply a venting of the suspicious emotion of revenge. As Ten asks, "how can

punishment, which involves the deliberate infliction of suffering or deprivation, be justified when it produces no good consequences, such as the deterrent, reformative, and incapacitative effects which utilitarians stress?"<sup>5</sup> It is important to stress that clause (2) of Pure Retribution closes escape routes that would claim that the harm produced by punishment is an unfortunate by-product of attempts to reform, deter, or whatever. The aim is to inflict harm.

As I said, to many Moral Restoration seems utterly mysterious. This is Sayre-McCord's point — how can the intentional infliction of harm for the sake of inflicting harm, "restore" a moral condition or "right a wrong"? As Sayre-McCord sees it:

punishment, so far as I can see, serves these purposes only by taking on a conventionally established symbolic role. Punishment works to "annul a crime" or "right a wrong" not naturally, so to speak, but only in a context within which suffering the intentional infliction of pain or harm of the sort in question is accepted as wiping the slate clean. Within the bounds set by general considerations of justice, precisely what can count as wiping the slate clean is a matter of achieving a shared understanding of what to accept. What might serve as proper payment for offence, like what might serve as legal tender, is largely a matter of what people collectively are willing to recognize and accept as such.

If this is right, then to the extent punishment can reasonably be seen as annulling the crime, righting the wrong, or reaffirming the victim's standing, reparations can as well, as long as it can take on the crucial symbolic role.

I shall argue that the Moral Restoration function is not simply a matter of conventional symbolism, but is deeply rooted in our concept of deserved punishment.

Lastly, even Political Possibility has been the subject of attacks. Pure Retribution is, some have claimed, impossible to implement by a public authority. As Ten says, although "there

is no complete agreement about what sorts of theories are retributive...all such theories try to establish an essential link between punishment and moral wrongdoing." If punishment gives people what they deserve, we need some way to calculate a person's overall moral deserts. But, says Sayre-McCord, we cannot really tally up a person's overall moral deservingness. Certainly the state, a blunt instrument, is unable to perform that task. We would have to know a person's overall deservingness throughout their lives to see what they really deserve — but who thinks that government could possibly do that? And even if it could, it is not clear to Sayre-McCord that it has a right to:

I am not saying that a state never rightly makes someone suffer. Assuming the state's coercive power is sometimes justified it will almost surely sometimes justifiably coerce in circumstances where someone is thereby made to suffer. What seems mistaken is the suggestion that the justification of such coercion will rest directly on the state's determination of moral defect. Making people suffer *because of their moral defects* is not legitimately within the state's purview. Moreover, to the extent one holds that those who are immoral deserve to suffer and that the state is properly involved in distributing that suffering in proportion the immorality, such a distribution of suffering would presumably have to take into account not just how immoral the criminal is, but also how much he has already suffered.

So even if the idea is not unacceptably inhumane and even if the mystery around it can be dispelled, it presents an impossible standard.

Let me turn, then, to my main task — defending the apparently inhuman, mysterious and impossible view that criminals deserve to be punished. We have seen why someone might wish to avoid this claim. But *how* might one go about rejecting this claim? It is, as John Kleinig points out, a very basic element of commonsense views about punishment. How do we get rid of so basic an idea as criminals deserve punishment? Well, the quickest and easiest route is to simply reject the entire idea of desert, as do many consequentialists and such radical reformers as Randy Barnett. What, then, is involved in rejecting desert?

There is a huge philosophical literature on the concept of desert, and I have no intention of trying to give a definite explication of the concept. Rather, I simply shall focus on what I think is the most important type of desert:

*Personal Earned Desert*: Alf deserves treatment *T* because of some past action *X* that Alf has performed.

Note two features of Personal Earned Desert. Most obviously, it supposes that the subjects of desert claims must be persons. We can, of course, intelligibly ascribe desert claims to objects as in "The book deserves to be read" or "Duluth deserves a better reputation." I, however, only wish to consider personal desert.<sup>9</sup>

Secondly, Personal Earned Desert supposes that the basis for the deserved treatment is some action that one has performed in the past. Again, there may be desert claims that are not about past actions, such as "the most beautiful entrant deserves to be Miss Universe." Although I cannot argue for it here, I think earned desert is the philosophically interesting and justifiable idea. It allows us, for example, to distinguish personal desert and merit. Consider the case of George the star law student who graduates at the top of his class but gets into a motorcycle

accident over the summer and is severely brain damaged. It seems he still deserves the award for the best law student: his past performance is what renders the treatment deserved *now*. However, he would not be selected on the basis of merit as the best person in a highly meritocractic law firm. Merit involves more forward-lookingness — it considers one's ability to do the job, not simply one's past accomplishments.

One reason for focusing on Personal Earned Desert is, I think, that many of the puzzles about the inhumanity of punishment derive from the puzzle about why we should doing something *now* because on what was done *in the past* even though no *good* will *come of it*.

Sayre-McCord approvingly cites Protagoras as objecting the past cannot be undone (which also concerns the problem of the mystery of punishment), and so our concern should be the future.

And if our concern is with the future, we will worry about merit and incentives, but not deserts.

So we might abandon all claims about Personal Earned Desert, and focus only on the future, not the past. But this is a radical move indeed. And, of course, it is not really an objection to punishment, but to the entire idea of personal earned desert. Although the consequentially-minded among us will not balk at this, it is radically revisionist. Studies of ordinary reasoners show that notions of deservingness are fundamental to their understandings of justice. As Mill observed, "it is universally considered just that each person should obtain that (whether good or evil) which he deserves; and unjust that he should obtain a good, or be made to undergo an evil, which he does not deserve. *This is, perhaps, the clearest and most emphatic form in which the idea of justice is conceived by the general mind*." <sup>10</sup> Perhaps we might be driven to radically revise our understanding of justice by dropping a core element. But surely, as moral reasoners, we have reason to explore understandings of justice and punishments that do not carry such a cost.

Assume for now, then, that we do not want to embrace a radically revisionist consequentialist conception that does away with desert as a basic moral category. On what other general grounds might one reject the idea that criminals deserve punishment? One line — perhaps the most important — of anti-punishment reasoning is motivated by what I shall call the *asymmetrical* conception of desert, according to which the concept of desert is asymmetrical insofar as people can deserve only positive, not negative, treatments. On this view, those who contribute to joint projects, for example, deserve rewards or benefits, but no one ever deserves the opposite of a benefit — a harm. If such an asymmetrical conception of desert could be defended one could, consistently, accept desert as a *bona fide* criterion of, say, economic justice insofar as a person can deserve to be rewarded for her productivity or effort but still reject any claim that criminals deserve to be punished.

I think, though, that the asymmetrical conception should be rejected. Although not as radically revisionist as completely rejecting the concept of deservingness, it is highly revisionist. It is certainly no part of commonsense morality that people can only deserve the good things in life. "Anyone who marries for good looks deserves all they get" by no means supposes that we can only deserve positive treatments. Of course some people do endorse the asymmetrical conception. I have quite a few colleagues who apparently believe that, while a student can deserve a good grade, it is impossible for anyone to deserve a bad one. Intuitively, I do not think this position is at all plausible: if we are going to have desert, I think it must be both positive and negative. However, we need not rely simply on intuitions. I wish to argue that our best theories of desert — our best theories as to what desert is all about — support what I shall call a

symmetrical conception of desert.

 $\mathbf{V}$ 

Before proceeding, let me explain what I mean by a symmetrical conception of desert. We can distinguish three levels at which the notion of desert is symmetric. The most obvious, and uncontroversial, symmetry claim is:

Concept-level Symmetry: The concept of desert is such that it involves both positive and negative desert claims.

John Stuart Mill, George Sher, Joel Feinberg<sup>11</sup> and others endorse this basic idea that there are both positive and negative sides to desert. More interesting — and I think seldom noticed — is what I shall call:

Theory-level Symmetry: On the best theory(ies) of desert, the same rationale that leads us to identify positive desert claims leads us to identify corresponding negative desert claims.

Here, the idea is that our best accounts of the nature of desert claims indicate that they are both positive and negative. I shall employ Theory-level Symmetry in what follows. I wish to stress that I do not advance the more radical idea of:

Claim-level Symmetry: For any positive desert claim D+ supposing conditions C, there is a corresponding negative desert claim D-, which also supposes conditions C.

If Claim-level Symmetry holds, all the considerations that are relevant in making a positive desert claim will also be relevant in a corresponding negative desert claim. I do not believe that claim level symmetry holds, certainly not without exception. A version of one case that has been put to me is Hero and Villain:

Hero. At great risk to his personal safety, Alf pulls three people out of a burning house. Later, he is awarded the Local Hero medal and a new BMW for doing such great good at such personal risk to himself. It is said he is deserving of the honor given the great good he did at risk to himself.

*Villain*: At no risk to his personal safety, Alf pushes three people into a burning house. Later, he is branded Local Menace and jailed for twenty years. It is said he is deserving of the stiff sentence because he was never in danger.

If Claim-level Symmetry held, all the considerations that are relevant in determining a person's positive desert would also affect considerations of negative desert. But in this case, while it seems that putting himself at risk was crucial in determining the deservingness of the Hero, it largely irrelevant in determining the deserts of the Villain, and that is why the Villain story seems so strange.

When I refer to the symmetry thesis, or the symmetrical conception of desert, I mean neither simply the true but rather bland concept-level symmetry, nor the very contentious claim-level symmetry, but *Theory-level Symmetry*.

#### $\mathbf{VI}$

Let us assume, then, that if we embrace desert as a basic moral notion, we embrace a symmetric conception of desert. However, even if we embrace a symmetric conception of desert, we still might avoid the conclusion that someone deserves to have harm inflicted on her — that is, punished. According to the *compensation theory of desert*, a person who has incurred a cost in achieving a social benefit, deserves to be compensated — the disutility incurred in bringing about the benefit should be wiped out. Nasty jobs, it is said, deserve extra pay to compensate

the workers for the disutility of the job. It would seem that on this view one does not deserve a benefit that makes you better off than when you started, but simply compensation for what was lost along the way. Adopting the compensation theory to negative desert, we could say that a criminal deserves to lose any improvement in his utility that resulted from his crime, which decreased the welfare of others or served to harm the common project. His gains should be annulled.

We can thus develop a symmetric compensation theory of desert:

Compensation Theory: Under conditions C, if Alf has performed act X which (1) is costly to him and (2) contributes to the common good or the success of a common project, then (3) Alf deserves treatment T, which compensates for that disutility. Under conditions  $C^*$ , if Alf has performed act Y which (1) benefited him and (2) detracts from the common good or the success of a common project, then (3) Alf deserves treatment  $T^*$ , which removes that additional utility.

We should note any theory of desert as compensation must identify the conditions that make desert claims relevant, such as voluntariness, the nature of the benefit or attack on the welfare of others, and so on. Hence conditions C and  $C^*$  must be specified. Because I have rejected Claim-level Symmetry, I do not suppose that C must be identical to  $C^*$ .

It is important to stress that it is no part of the compensation theory of negative desert that the perpetrator should compensate the victim: that is a separate matter. What is crucial is that the criminal loses any gain achieved through his trespasses. Note that this falls short of endorsing Pure Retribution. According to Pure Retribution the criminal deserves to have harm inflicted on her and in that way be made worse off. But negative compensation simply aims that the criminal be no better off. It is for this reason that I said one might embrace the idea of negative desert and

have still plausibly reject the notion of punishment.

Still, I think this is cold comfort for the anti-retributivist for two reasons. First, the compensation theory provides a powerful rationale for a symmetric conception of desert. The heart of the compensation theory is an ideal normative baseline. The arguments that indicate that contributors should not fall below the baseline because of their contributions also justify the claim that detractors should not rise above their baseline because of their transgressions. The justifications for the endorsing positive desert are equally well suited to justifying negative desert. Second, while we do not quite have Pure Retribution, we are being led to the justification of treatments of criminals that seek to deprive them of what they have —which explicitly aims at detracting from their welfare. It may not be punishment understood as intentional harm, but is it awfully close.

However, I think the compensation theory is ultimately inadequate as a theory both of positive and negative desert. It is inadequate as a theory of positive desert insofar as it never deems that a contributor deserves to be better off than a non-contributor: she simply deserves to have her additional disutility removed so that she ends up no worse off. That seems a very stingy theory of positive deserts. It is inadequate as an account of punishment as the theory presupposes that the criminal obtain gains, which then can be wiped out. But if the criminal gained nothing — if, as is so often the case, her criminality was a disaster for her as well as for her victims — then there appears no grounds for punishment. If you haven't gained anything, your gains cannot be taken away. To be sure, some philosophers who have thought along these lines have argued that the mere breaking of the law is an unfair advantage — a sort of illicit gain — which must be annulled. But if, as is usually the case, criminals are their own worst enemy, they do a pretty good job at annulling their own gains.

#### VII

Let us then consider a more fruitful way of explicating desert:

Reward Theory: Under conditions C, a person who has X-ed, and so has contributed to the common good or a common project deserves treatment T as a reward — she deserves a beneficial treatment. Under conditions  $C^*$ , a person who has Y-ed, and so has subtracted from the common good, or a common project, deserves a treatment  $T^*$ , a disreward (a punishment) — she deserves a harmful treatment.

This, of course, is a symmetric theory of reward. The core idea is that while certain sorts of past contributions make appropriate positive treatments of contributors, this same understanding of appropriate treatment provides a rational for interest-detracting treatments of discontributors. I am not claiming that the negative notion logically follows from the positive, or that one is caught in a contradiction to affirm the negative and deny the positive. My claim, rather, is that the desert-based reasoning that makes positive treatments appropriate also makes negative treatments appropriate. This is especially clear in expressive theories of desert. On an analysis such as Christi Dawn Favor's, 13 the appropriateness of the treatment stems from expressions of attitudes to various actions; the positive treatment is an expressive approval. On her view, the same considerations that lead us to express favorable attitudes to certain actions — say, contributing to the collective endeavor — also commend negative expressions towards those who undermine the common endeavor.

The Reward Theory, then, provides a strong case for a symmetric conception of desert.

Although I shall not examine other current theories, let me simply observe that other important accounts also seem symmetric or bivalent in this way. If positive desert treatments are explicated

in terms of a fittingness relation — if it is fitting to treat positively contributors to the common good — then that same fittingness relation would seem to call for negative treatments for detractors to the common good. Again, if (as I have argued) producers deserve to control the fruits of their labors, those who attack production (say by invading others) would seem to deserve to loose the fruits of their labor. Is

Note that if we accept the symmetric conception of desert, we have arrived at a justification for Pure Retribution: treatments that set out to harm people. The very same relation — compensation, <sup>16</sup> reward, fittingness, labor-production — that grounds a positive, beneficial, desert treatment also grounds a negative, harmful, desert treatment — that is, punishment.

#### VIII

Having briefly sketched the idea of, and case for, a symmetric conception of desert, let us return to our three worries about retribution. Now recall that Sayre-McCord argues that if we could come up with some theoretical justification of Pure Retribution — even if we could make some sense of the claim that people deserve to be harmed — inflicting such harm could not be within the state's competency. This is the impossibility worry. As he sees it, punishment is inflicted because of one's moral defects, and the state is not competent to evaluate people's moral defects. Let's see if we can better understand the competency issue by focusing on positive desert, and then go back to the always-troubling case of the negative.

It has been widely, though of course not universally, held that the government has some role in ensuring that those who produce and labor get what they deserve. According to many advocates of a private property market economy, it has the crucial virtue of rewarding the industrious and the productive. On this view, helping to ensure that people get what they

deserve requires that private property and competition be protected.<sup>17</sup> In contrast, many advocates of the welfare state — from L.T. Hobhouse to David Miller<sup>18</sup> — have insisted that some form of government-sponsored redistribution is required to help ensure that those who contribute are adequately rewarded. On these very different views of government, although government cannot ensure that holdings perfectly track desert, it can take significant action to help ensure that people get what they deserve. To be sure, some such as F.A. Hayek and John Rawls<sup>19</sup> accord government no significant role in this regard; but it is not especially controversial to claim that government is competent over areas of positive, economic, desert.

Note that while on this account of positive desert producers have a moral claim to be rewarded for their contributions, this is *not* a claim about their overall *moral* deservingness. This is a crucial point. Consider the contrast between a moral desert claim and a claim about one's moral deserts:

A: Moral Desert Claim. Because of one's past action X, one now has a moral claim to treatment T, which one deserves.

*B: Claim to Moral Deserts.* Because of one's past moral merits/demerits, one now deserves treatment *T*.

It seems to me quite right that the state is not competent to give us, B, our moral deserts. Either in a positive or negative sense, it seems the impossibility worry is perfectly reasonable about claim B.

This has been a real source of confusion in the literature: again and again it is asserted that Pure Retribution presupposes B. We are told by a wide range of theorists that retributive punishment is about giving people their moral deserts. But it is puzzling what that should be the case with criminal punishment. The relation between law and morality is complex, and very few

political theories simply depict the law as expressing all the demands of morality. Now, if the law does not simply express the demands of morality, why should we suppose that the legal punishment is supposed to reflect the criminal's overall moral worth? If, however, we understand the law as providing, crucially, the framework for social cooperation and peace, then we will see the law as concerned with those who attack the basis of peace and cooperation: those anticontributors who detract from, rather than contribute to, civil society.

There still is a moral dimension to this: it seems right that people have a moral claim to be rewarded for their contributions, just as they have a moral liability to be punished. But to explain that all we need is A, a moral claim to get what one deserves. And claim A does not seem impossible: it is certainly presupposed by a host of positive desert claims.

There is an interesting anti-libertarian twist to this argument. Many libertarians argue that it is impossible for the state to seek to give people what they positively deserve, yet they are willing to accept a retributive theory according to which criminals deserve punishment.<sup>20</sup> If I am right here, the symmetry thesis suggests they face the mirror-image of the problem confronted by social democrats who think the state can reward economic deserts but are appalled at retribution. Of course they may show why the state's competency is so asymmetric, but on the face of it seems puzzling that it should be so. It is, in the end, a similar sort of issue.

Focusing on desert as a symmetric concept, then, allows us to better understand the sort of negative desert claims that are involved in punishment, and helps us to avoid what seems to me the common error of supposing we need claim B, rather than the much weaker claim A.

#### IX

I have tried to sketch here an argument — and of course in this context it can be no more than a

sketch — to show how, once we see the symmetric nature of desert, we grasp better the sort of negative desert claim that is relevant, and the impossibility worry dispels. Or if it not dispelled, it is transferred to a set of worries about the general idea of government giving anyone what they deserve, including economic deserts.

Let me turn briefly to the "mysteriousness" worry concerning the Moral Restoration claim. How can punishing a person help square things away with them? Again, consider first positive deserts. Suppose an initial situation in which people are free to decide whether or not to contribute to some joint project. All rights and duties are honored, all claims are met. Some people contribute, other's don't. The contributors advance the project: we are better off for their contributions. It is plausible in this sort of case, I think, to claim that the contributors deserve benefits. Suppose we accept that they do deserve a reward, and treat them in a beneficial way They get what they deserve. Is it mysterious to say that we have squared things away with them, or at least have helped to do so? I don't see how — we have given them what they deserve, and we have met their claim on us. We have helped return to the initial status quo in which all rights and obligations are met. In Aristotle's terms, we have restored equality. <sup>21</sup> Pace Sayre-McCord it is not simply a matter of symbolic, conventional, expression that awarding our contributors what they deserve squares things away with them. Sayre-McCord is, of course, right that conventions may be employed, as for example concerning the currency we use, and all the conventions that relies on. But that squaring away requires that we honor their claim to positive treatments is not simply conventional: it is intrinsic to just treatment.

Turn, then, to the negative case. If we square away things when we give the enhancer what she deserves, why should it be mysterious to claim that we help square away things when we give the detractor what he deserves? Why is it the case that only benefits can serve to square

things away? As long as we accept a symmetrical principle of desert, the idea that negative deserts can square things away is no more peculiar than the claim that giving positive deserts can square things away. Again, Sayre-McCord is right that conventions may determine what negative treatments are meted out. It is surely a mistake of retributivists to claim — as they have sometimes done — that the details of what constitutes a just punishment can be read off the very concept of retribution. My Pure Retribution claim, though interestingly strong, does not make any such (in my view) wild claim.

I do not assert that there is no mystery here. There is a general mystery about debts and squaring things away. How can it be that, as Personal Earned Desert claims, what a person has done in the past requires us now to treat him in a certain way? The past determines the present, not by a causal connection, but by the crux of desert claims: past actions give us reasons to treat the actors in certain ways now. On most standard, instrumentalist-inspired, theories of rationality, it is hard to see how the past can generate reasons in this way. If we are not seeking a future satisfaction or value, how can we have reason to engage in a certain type of treatment? There are some interesting mysteries here, but I do not think they are mysteries about punishment. They are perfectly general worries about the rationality of desert claims. In order to resolve these mysteries we cannot simply focus on narrow issues about punishment and retribution: they are much broader, and lead to much more general philosophical puzzles about the nature of practical reason and non-instrumental action.

 $\mathbf{X}$ 

I have thus far considered the mysteriousness of desert squaring things away, and the impossibility worries. What of the inhumanity and barbarism worry about Pure Retribution?

Surely, we might think, this is about only negative, not positive, desert. Yet here too, I suggest, the worry is better understood as one about desert, not about punishment. It is the essence of Personal Earned Desert that it ties present treatment to an agent's past actions: we do things to her now just because of what she did in the past. In particular, we do not treat her in the way that we think will do her or the world good, or in a way that makes it or her better. Our reason for the treatment lies in the past. This means that we have strong reason to award a person her deserts regardless of whether it does good or ill to the world or to her.

This is masked in positive desert claims. Because we are giving a person a treatment that constitutes a reward, and so in a narrow sense a benefit, it looks as if we are doing good to her: and thus the humanitarian or welfarist impulse does not apparently clash with desert. But if there is no clash, it is only fortuitous; for the good or ill that will result over the long run to the agent by giving her the deserved benefit is not ultimately relevant. On American cable television — the VH1 channel — there is a resolutely depressing program called "Behind the Music." It shows, episode after episode, how rock stars, who sought for years after success, and many of whom deserved it, are typically ruined by it. The fruits of their labors give them enough money to buy enough drugs to ruin themselves. Getting the positive treatment you deserve is very often the thing that ruins your life. But that doesn't mean that you do not deserve it. Although you deserve a beneficial treatment, it is not claimed that this treatment ultimately is better for you than not getting it — we often do not know whether it turns out to be good or bad, but it is still what you deserved.

The tension between humanitarian welfarism and desert is simply clearer with negative desert. Here, desert aims to harm, and so the conflict is more obvious. Of course here too, it may actually help: a few people probably actually benefit from being punished. If a criminal says

"Going to jail turned my life around, and was the best thing that could have happened to me," we do not say "Oops! We didn't really punish you then!" This does not show that she was not punished, for a harmful treatment was indeed inflicted in her. The basic insensitivity to the good or ill that comes out of honoring the claim is, I think, a more general feature of desert thinking. Any worries about it should be located there.

#### ΧI

Surely, it might be said, there is one crucial asymmetry between positive and negative desert: whereas positive desert, because it is constituted by a beneficial treatment, requires no special justification, negative desert, because it inflicts a harm, always requires a justification. Inflicting harms on others always requires special justification, while showering them with benefits does not.

Now this may be accepted: at the level of all things considered judgments, the retributivist may accept that punishing people faces greater barriers than positively rewarding them. From the perspective of desert claims, we might say, positive and negative treatments are on par, but because of the *prima facie* wrongness of inflicting harm, there is a greater barrier to justifiably giving people their negative deserts. That we should be careful about giving people their negative deserts, because unjustifiable harming is itself a wrong, does not show that criminals do not deserve punishment, but it may provide a powerful reason why punishment must be surrounded by safeguards and procedures.

But, it might be pressed, isn't it *always* wrong to harm others, and so is not punishment always wrong? It seems not. John Stuart Mill's famous harm principles does not assert that harm is always wrong, but rather than harming always requires justification.<sup>22</sup> Moreover, if we take the

view suggested by John Locke, that a criminal forfeits some or all of his rights (not to have certain negative treatments inflicted on her), then it is not even the case that the criminal rights not to be harmed are overridden by punishment, for she has forfeited these rights.<sup>23</sup>

#### XII

I have tried to do two things in this paper. First, I tried to answer Sayre-McCord's worries about desert and retribution by sketching a defense of the Pure Retribution, Moral Restoration and Political Possibility claims, by showing that they are endorsed by the Reward Theory, understood as a symmetrical theory of desert. And, I have argued, something approaching Pure Retribution is endorsed by the Compensation Theory. Although I cannot argue for it here, I have indicated that they are also endorsed by other plausible theories of desert. I want to stress that my case does not depend on any specific theory of desert:<sup>24</sup> I believe that my thesis holds for a range of theories of deservingness. In this sense I think my claims are fairly robust: the symmetry thesis holds for an interestingly large class of accounts of deservingness.<sup>25</sup>

Note that I have not simply dismissed or dispelled all the worries about punishment with which I began. To some extent, I think the impossibility worry can be dismissed: it supposes that we are seeking to give a person her overall moral deserts, when I think no such claim is plausibly made in justified systems of social punishment. But the other two worries — the inhumanitarian and mysteriousness worries — I have tried to relocate. They are genuine worries, but I have tried to suggest that they are worries about the notion of desert, and its relation to the dominant theory of practical rationality and related welfarist consequentialism. I do not think we can resolve these worries with more work on the notion of punishment.

#### **NOTES**

<sup>&</sup>lt;sup>1</sup> Gerald Gaus, "Does Compensation Restore Equality?" in John W. Chapman, ed., *Nomos XXXIII: Compensatory Justice*. (New York: New York University Press, 1991), pp. 45-81.

<sup>&</sup>lt;sup>2</sup> C. L. Ten, *Crime, Guilt, and Punishment* (Oxford: Clarendon Press, 1987), p. 46.

<sup>&</sup>lt;sup>3</sup> Immanuel Kant quoted in George Sher, *Desert* (Princeton, N.J.: Princeton University Press, 1987), p. 70.

<sup>&</sup>lt;sup>4</sup> Thomas Hobbes, *Leviathan*, Michael Oakeshott, ed., (Oxford: Blackwell, 1948), p. 202 (ch. 28).

<sup>&</sup>lt;sup>5</sup> Ten, Crime, Guilt and Punishment, p. 46.

<sup>&</sup>lt;sup>6</sup> Ibid., p. 38.

<sup>&</sup>lt;sup>7</sup> John Kleinig, *Punishment and Desert* (The Hague: Martinus Nijhoff, 1973).

<sup>&</sup>lt;sup>8</sup> Randy E. Barnett, *The Structure of Liberty* (Oxford: Oxford university Press, 1988), p. 319.

<sup>&</sup>lt;sup>9</sup> I am indebted here to my former colleague, Christi Dawn Favor.

<sup>&</sup>lt;sup>10</sup> John Stuart Mill, *Utilitarianism* in John Gray, ed., *On Liberty and Other Essays* (Oxford: Oxford University Press, 1991), ch. V. Emphasis added. On the attitudes of ordinary reasoners to desert, see my *Social Philosophy* (Armonk NY: M.E. Sharpe, 1999), pp. 102ff.

<sup>&</sup>lt;sup>11</sup> For Mill see *Utilitarianism* ch. V; for Sher see *Desert*, p. 4; for Feinberg see *Doing and Deserving* (Princeton, N.J., Princeton University Press, 1970), p. 67.

<sup>&</sup>lt;sup>12</sup>See David Miller *Social Justice* (Oxford: Clarendon Press, 1976), p. 103 and Julian Lamont, "Productivity, Compensation, and Voluntariness" in Gerald F. Gaus, Julian Lamont and Christi Dawn Favor, eds., *Values, Justice and Economics* (Buffalo, NY: Humanity Books, 2001).

<sup>&</sup>lt;sup>13</sup> Christi Dawn Favor, "Expressive Desert and Deserving Compensation" in Values, Justice and Economics.

<sup>&</sup>lt;sup>14</sup> See Geoffrey Cupit, *Justice as Fittingness* (Oxford: Clarendon Press, 1996).

<sup>&</sup>lt;sup>15</sup> See my *Value and Justification* (Cambridge: Cambridge University Press, 1990), pp. 410ff.

<sup>&</sup>lt;sup>16</sup> Subject to the qualification in section VI.

<sup>&</sup>lt;sup>17</sup> I have argued this in *Value and Justification*, §26.2.

<sup>&</sup>lt;sup>18</sup> See Hobhouse, *The Elements of Social Justice* (London: Allen and Unwin, 1949); Miller, *Social Justice*.

<sup>&</sup>lt;sup>19</sup> Hayek dismisses the very idea of social justice, emphasizing the classical liberal position that there is no such thing as "value to society." "Services can only have value to particular people (or an organization), and any

particular service will have different value to different members of the same society." F.A. Hayek, *Law Legislation and Liberty*, vol. 2, *The Mirage of Social Justice* (Chicago: University of Chicago Press, 1976), p. 75. For Rawls's position on desert, see *A Theory of Justice*, revised edition (Cambridge, MA: Harvard University Press, 1999), pp. 88-89, 273-277.

<sup>&</sup>lt;sup>20</sup> Not Randy Barnett. See above note 8.

<sup>&</sup>lt;sup>21</sup> See my "Does Compensation Restore Equality?"

<sup>&</sup>lt;sup>22</sup> See my *Social Philosophy*, ch. 8.

<sup>&</sup>lt;sup>23</sup> John Locke, *Second Treatise of Government* in Peter Laslett, ed., *Two Treatises of Government* (Cambridge: Cambridge University Press, 1960), §§82, 172.

<sup>&</sup>lt;sup>24</sup> In fact, I have myself criticized the Reward Theory. See *Value and Justification*, Appendix B.

<sup>&</sup>lt;sup>25</sup> On the virtues of robustness in political theory, see my *Justificatory Liberalism* (Oxford: Oxford University Press, 1996), ch. 1.