

# General Aspects of Law

## GALA

DEAN'S SEMINAR ROOM

(RM. 215B)

THURSDAY

**FEBRUARY 26, 2009 - 4:00 P.M.**

---

***“DESERT-CONSTRAINED PLURALISTIC  
CONSEQUENTIALISM:  
TRIUMPH AND CHALLENGES”***

By

**MITCHELL N. BERMAN**

**PROFESSOR OF LAW**

**RICHARD DALE ENDOWED CHAIR**

**UNIVERSITY OF TEXAS LAW**

NOTE

Discussion will proceed without a presentation of the paper by the author. Papers are distributed to those on our mailing list or electronically when available at the GALA website:

< <http://www.law.berkeley.edu/1151.htm>>.

Others may obtain copies from Ms. Amatullah Alaji-Sabrie at 510.642.3627 or [asabrie@law.berkeley.edu](mailto:asabrie@law.berkeley.edu). Copies can be made available in an alternate format upon request. GALA events are wheelchair accessible. For any disability-related accommodations advance notice is requested.

16 February 2009

To Participants at the GALA Workshop:

The attached paper is a very tentative first draft of an essay to be contributed to a forthcoming edited volume entitled "The Philosophical Foundations of Criminal Law." The aim of that book (you might find it useful to know) is "to identify and illuminate a core set of foundational issues in the philosophy of criminal law" and to "help to set the terms and the direction of philosophically informed criminal law theory for the next decade or more."

As you will see, the thrust of my essay is to question the conventional division of punishment theories between retributivism and consequentialism. You will also see that the essay is incomplete, rough and, in places, impressionistic. I am extremely eager for your comments, criticisms, and suggestions.

Thank you in advance for your engagement with the paper.

Mitch

DESERT-CONSTRAINED PLURALISTIC CONSEQUENTIALISM:  
TRIUMPH AND CHALLENGES

Mitchell N. Berman\*

INTRODUCTION

The philosophy of criminal law covers a broad range of concerns. Its practitioners explore such diverse conceptual and normative matters as the character of a culpable act and the proper contours of criminal liability for an omission, the principles of causation, the difference between defenses of justification and of excuse, the nature of complicity, and a variety of puzzles involving attempts, among innumerable other topics. Historically, however, one question has dominated the rest: in virtue of what is the state morally justified in subjecting an individual to criminal punishment—i.e., the intentional infliction of suffering and/or the deprivation of substantial liberties, joined to moral censure or condemnation? Answers to this question routinely travel under the heading of “theories of punishment,” though “justifications for punishment” would be more apt. Refining, defending, and critiquing theories of punishment have been the central concerns of philosophers of the criminal law. For centuries it has been a vigorous and fractious debate.

In the early years of the 21<sup>st</sup> century, however, criminal law theorists have reached a degree of consensus regarding an answer to this central question that is remarkable by historical standards. The theory of criminal punishment on which they are coalescing has three principal elements. To a first approximation, they are: first, that punishment is all-things-considered morally justified, if at all, by the net good consequences we reasonably expect it to produce; second, that the good consequences at which punishment aims are varied, and may include (though it need not) the state of affairs in which a wrongdoer suffers, or experiences deprivation, in proportion to and on account of his blameworthy wrongdoing; and third, that

---

\* Richard Dale Endowed Chair in Law, the University of Texas at Austin.

pursuit of net good consequences is properly constrained (though possibly not absolutely) by the moral command that the state not knowingly inflict suffering absent, or in excess of, the ill-desert of the person punished. I will call this account “desert-constrained pluralistic consequentialism” (“DCPC”).

This essay has two goals. Principally, it aims to substantiate that desert-constrained pluralistic consequentialism is as widely endorsed as I have just claimed. In doing so, it will elaborate on the theory, situate it within an historical context, and address anticipated objections. Secondly, it identifies several questions that warrant attention from the community of philosophers of criminal law who endorse DCPC and yet are presently insufficiently theorized. At the outset, though, one caveat: In drawing attention to the present-day “triumph” of desert-constrained pluralistic consequentialism, I do not mean to suggest that this state of theoretical agreement will be permanent. I am not announcing “the end of history” in the domain of the philosophy of punishment. I aim only to present a picture of the state of Anglo-American criminal theory at a moment in time.

## I. TRIUMPH

### *A. E Pluribus Duo.*

Philosophers writing in the Western tradition have endeavored to justify criminal punishment against moral objections since the Greeks, even if, from today’s perspective, the dominant non-contemporary figures date no farther back than to the mid-eighteenth century. Viewed through one lens, this lengthy tradition has produced a rich diversity of justificatory theories, including deterrence (Bentham and Beccaria), reform (Plato), retribution (Kant), annulment (Hegel), and denunciation (Durkheim).

A striking feature of twentieth century punishment theory, however, has been the steady and generally successful pressure to fold this seeming multiplicity of justifications into a simple dichotomy of justifications that closely mirrors the fundamental organizing distinction in moral theory between consequentialism and deontology. Thus have commentators routinely insisted that deterrent, reform, and denunciatory, expressive, or educative theories are most perspicuously understood simply as emphasizing

different—but not incompatible—mechanisms by which punishment brings about a varied lot of desirable consequences, while Kantian, Hegelian, and similar theories are best viewed as arguing that punishment is right or fitting in itself. To be sure, this effort at binary classification between what are often termed “forward-looking” and “backward-looking” theories has always met with some resistance. But a central strand in the story of the philosophy of criminal law over the past two centuries consists of the gradual refining of a somewhat more diverse array of competing theories of punishment into the two dominant traditions generally recognized today: the consequentialist tradition tracing its roots back to Beccaria and Bentham, and the retributive tradition that claims Kant as its patron saint.

*B. And Then There Was One*

That is the ruthlessly abbreviated thumbnail sketch of two hundred or more years of criminal law theory. The thesis of this essay is that the past two decades have witnessed a steady convergence from two theories—the consequentialist theory that justify punishment in forward-looking fashion by reference to the net good consequences it can be expected to produce through a variety of causal mechanisms, and retributivist theories that justify punishment in strictly backward-looking fashion by reference, paradigmatically, to the offender’s desert—to a single pluralistic consequentialism. This convergence between consequentialism and retributivism is the product of accommodation from each side toward the other.

### 1. Refining Retributivism

Even as theorists sought to divide the universe of punishment justifications in two, a concise statement of retributivism remained notoriously elusive. In a well-known article from 30 years ago, for example, John Cottingham distinguished nine distinct theories that had been classified, by its proponents or others, as retributivist. These included the theses: that offenders deserve to be punished; that through punishment a wrongdoer repays his debt to society; that punishment annuls crime; that punishment restores conditions of fair play between offenders and the law-abiding; that punishment absolves a society's blood guilt for crime; and that punishment satisfies the longing of victims and the public for justice and revenge.<sup>1</sup> Although Cottingham rightly denied that several of these merited the retributivist label, he did not propose his own statement of what retributivism is, or maintains. That awaited Michael Moore's influential work in the early 1980s. According to Moore,

Retributivism is a very straightforward theory of punishment: We are justified in punishing because and only because offenders deserve it. Moral responsibility ("desert") in such a view is not only necessary for justified punishment, it is also sufficient. Such sufficiency of justification gives society more than merely a *right* to punish culpable offenders. . . . For a retributivist, the moral responsibility of an offender also gives society the *duty* to punish. Retributivism, in other words, is truly a theory of justice such that, if it is true, we have an obligation to set up institutions so that retribution is achieved.<sup>2</sup>

Moore is, unquestionably and by a wide margin, the most prominent contemporary retributivist, and this formulation has been widely seen to capture the essence of contemporary retributivism. That is not to say, however, that contemporary retributivists are apt to find this statement entirely acceptable, or even entirely clear. Let us put aside, for the moment, Moore's claims regarding the "duty" or "obligation" to punish (we will pick them up again in Part II), and focus on the notion of desert. As a first pass, the core retributivist claim would hold that the ill desert of an offender is a necessary and sufficient condition for the state to punish him. But it remains to identify with some greater precision just what it is that offenders deserve. Moore's

---

<sup>1</sup> John Cottingham, *Varieties of Retribution*, 29 *Phil. Q.* 238 (1979).

<sup>2</sup> MICHAEL MOORE, *LAW AND PSYCHIATRY: RETHINKING THE RELATIONSHIP* \_\_\_ (1984).

formulation suggests that what they deserve is punishment. Yet if we accept this premise, many commentators have complained, the argument for retributivism appears close to tautological. To avoid the charge of tautology, many retributivists would endorse this slight revision: what offenders deserve is to experience suffering, or deprivation, and punishment is the means of bringing about this deserved state. More precisely still: wrongdoers deserve to suffer (i.e., to endure some negative experiential state) on account of, and in proportion to, their blameworthy wrongdoing. Let us call this the *desert claim*.

Because the very notion of desert is at least a touch mysterious, many retributivists have translated the desert claim into the language of intrinsic goodness. Here's a candidate formulation of what I will call the *retributive intrinsic good claim* (or often the *intrinsic good claim*, for short): It is intrinsically good (or intrinsically valuable) that one who has engaged in wrongdoing suffer on account of, and in proportion to, his blameworthy wrongdoing. In preferring the *intrinsic good claim* to the *desert claim*, theorists have supposed that the two are equivalent.<sup>3</sup> In 1993, Moore himself endorsed the intrinsic good claim. Indeed, he allowed that the intrinsic good claim captures the core meaning of retributivism: "what is distinctively retributivist is the view that the guilty receiving their just deserts is an intrinsic good."<sup>4</sup>

As David Dolinko was among the first to recognize, acceptance of the intrinsic good claim by the foremost contemporary retributivist poses a profound challenge to our dominant classificatory scheme.<sup>5</sup> For while retributivists might believe that the realization of deserved suffering is an intrinsic good, they must not believe that it is the *only* intrinsic good, for it would be implausible to insist

That the overall goodness of any state of affairs depends *exclusively* on how much punishment-of-guilty persons it contains, regardless of whatever else that state of affairs contains. . . . Indeed, to insist that *only* the quantity of "the guilty receiving punishment" affects the goodness of a state of affairs implies the absurd conclusion that a state of affairs wherein no one ever commits any crime at all lacks goodness altogether!<sup>6</sup>

---

<sup>3</sup> See, e.g., Davis; Hurka; Ewing.

<sup>4</sup> Michael S. Moore, *Justifying Retributivism*, 27 *Israel L. Rev.* 15, 19 (1993) (emphasis omitted).

<sup>5</sup> David Dolinko, *Retributivism, Consequentialism, and the Intrinsic Goodness of Punishment*, 16 *L. & Phil.* 507 (1997).

<sup>6</sup> *Id.* at 513-14.

That would be an absurd view, and Moore explicitly disavows it.<sup>7</sup> But then it might seem that retributivism is not interestingly different from other consequentialist approaches to punishment, most of which also recognize that the consequences the realization or pursuit of which justifies the imposition of criminal punishment are of a varied sort. After all, the principal reason for terming the alternative to retributivism “consequentialism” rather than “utilitarianism” is precisely to deny that a consequentialist about punishment must believe that utility constitutes the only metric of value. If consequentialist justifications of punishment are, as a class, pluralist about value, then they comfortably encompass intrinsic-good retributivism.

Now, both Moore and Dolinko deny, albeit in different ways, that retributivism is best viewed simply as a consequentialist theory that isolates deserved suffering as a particularly weighty value. Moore denies that retributivism *need* be assimilated to consequentialism: in his view, there are consequentialist and deontological forms of retributivism, both of which are sound. Dolinko denies that retributivism *can* be assimilated to consequentialism. We will consider these objections in Part II. For now, let us accept Dolinko’s initial observation as our provisional conclusion: embrace of the intrinsic good claim, by Moore and other retributivists, converts retributivism into a consequentialist justification for punishment.

## *2. Constraining Consequentialism*

Recall that Moore had claimed that an offender’s ill-desert constitutes both a necessary and sufficient reason for the state to punish him. His embrace of the intrinsic good claim along with his recognition of a multiplicity of intrinsic goods demands that his contention that bringing about the good of deserved suffering is a sufficient condition for punishment be significantly qualified or at least clarified. If retributivists, like most of us, recognize a plurality of intrinsic goods then they are apt to recognize a plurality of intrinsic bads as well. It would seem to follow, then, that punishment might not be justified all things considered in a particular case if the bads it would produce outweigh the goods, including the good of realizing deserved suffering.

Accordingly, it seems untrue that realizing deserved suffering can be a sufficient condition of justified

---

<sup>7</sup> See, e.g., Moore, *Justifying Retributivism*, at 34 (“It would be a crude caricature of the retributivist to make him monomaniacally focused on the achievement of retributive justice. The retributivist like anyone else can admit that there are other intrinsic goods, such as the goods protected by the rights to life, liberty, and bodily integrity.”).

punishment. In fact, Moore grants precisely this. Any talk of “sufficient conditions,” he notes, is context-sensitive:

Within the set of conditions constituting intelligible reasons to punish, the retributivist asserts, desert is sufficient, i.e., no other of these conditions is necessary. Of course, other conditions outside the set of conditions constituting intelligible reasons to punish may also be necessary to a just punishment, such as the condition that the punishment not violate any non-forfeited rights of an offender.<sup>8</sup>

Put otherwise and more generally, Moore means to assert only that the realizing of deserved suffering supplies a sufficient reason to punish in the absence of overriding reasons not to punish. If the sufficiency condition is understood in this way, it is entirely consistent with consequentialist theories of punishment.

The assertion that the realization of deserved suffering constitutes a necessary condition for just punishment stands on different footing. To start, it warrants emphasis that the “necessary condition” claim is more modest than it might first appear. The claim is neither that an instance of punishment is necessarily unjust if it inflicts undeserved suffering despite the punisher’s genuine belief that the punishment is deserved, nor that a punishment practice or institution is unjust if it foreseeably results in undeserved punishment. The necessity condition maintains only that it is unjust to inflict punishment on someone known (or believed?) not to deserve the suffering imposed. Even thus reframed, however, the retributivist claim that the infliction of deserved suffering is a necessary condition of just punishment puts retributivism at potential odds with consequentialist justifications for punishment as the retributivist sufficiency condition did not, for full-blooded consequentialists reject the necessity claim even in this qualified form.

But—and here’s how consequentialism has moved in recent years to meet retributivism half way—few if any contemporary consequentialists about punishment are full-blooded. Consequentialist movement toward acceptance of retributivism’s necessity condition is customarily traced to the mixed theories Rawls and Hart advanced half a century ago. In Rawls’s rule-utilitarian picture, legislators justify criminal justice institutions and practices on consequentialist grounds, while judges justify the punishment of individual

---

<sup>8</sup> Moore, *Justifying Retributivism* at 35.

offenders on the non-consequential ground that he or she violated a legal command.<sup>9</sup> Similarly, Hart described the “general justifying aim” of the institution of punishment as crime reduction, but argued that pursuit of this consequentialist goal is constrained by a principle of “retribution in distribution” that permits imposition of punishment only on “an offender for an offense.”<sup>10</sup>

Notoriously, however, one can violate the terms of an offense without being morally blameworthy and therefore without incurring moral ill-desert. And neither Rawls nor Hart insisted, as a retributivist would, that justice demands that punishment not be imposed on someone known not to deserve the suffering imposed—a moral view that John Mackie would soon dub “negative” retributivism,<sup>11</sup> and that Antony Duff would call, perhaps even more aptly, “side-constrained consequentialism.”<sup>12</sup> Today, self-described consequentialists have gone beyond Rawls and Hart to routinely accept that it is impermissible to knowingly punish the innocent, and that even unknowing punishment of the innocent is a very considerable bad.<sup>13</sup> I will call a consequentialist theory that is limited in this way “desert-constrained consequentialism” to foreground clearly the nature of the side-constraint—namely, knowingly punishing absent or in excess of ill-desert.

### 3. Summary

By small steps, self-described retributivists and self-described consequentialists have nearly converged on a theory of the justifiability of criminal punishment that contains the following three elements: (1) the institution of criminal punishment, and the infliction of punishment in a given case, are morally justified by the good consequences sought to be realized; (2) the states of affairs that are intrinsically good, and that therefore can contribute to the justifiability of punishment, are varied and dependent upon one’s particular theory of value; and (3) it is at least presumptively impermissible to knowingly punish those who lack ill-desert.

---

<sup>9</sup> John Rawls, *Two Concepts of Rules*, 64 *Philosophical Review* 3 (1955).

<sup>10</sup> H.L.A. Hart, *Punishment and Responsibility* (1968), pp. 8-12.

<sup>11</sup> J.L. Mackie, “Morality and the Retributive Emotions,” *Criminal Justice Ethics* 1 (1982): 3-10, p. 3.

<sup>12</sup> R.A. Duff, *Punishment, Communication, and Community* (Oxford: Oxford University Press 2001), p. 11.

<sup>13</sup> See, e.g., Husak.

This does not mean that all theorists agree on the proper shape of the criminal law. Not at all. Of the potentially limitless points of disagreement, at least two deserve mention here. First, as already emphasized, theorists differ on their theories of value. To take the most prominent example, pluralistic consequentialists about punishment differ regarding whether it is a good at all, and if is a good, of what weight, that wrongdoers suffer on account of their wrongdoing. The convergence thesis does not deny this difference. Rather, it downplays this difference on the grounds that it is more salient than significant. That is to say, if we divide theories of punishment by reference to the particular types of goods that the theory invokes, or heavily relies upon, as potential justification for punishment, we would not be left with two theories—call them “retributive consequentialism” and “non-retributive consequentialism”—rather than one, but dozens or hundreds rather than two, for every different axiological position would constitute its own punishment theory. (For example, some consequentialists about punishment who deny that suffering is ever good might nonetheless be deontologists about morality and therefore believe that causing harm to others in ways that violate deontological moral commands is worse than causing identical harms that do not violate deontological moral commands, and therefore that a reduction in wrongful harms is of greater intrinsic value than the same reduction in non-wrongful harms, all else equal.) To avoid an unwieldy diversity of punishment theories, I believe that theories of punishment are more illuminatingly classified at a level of generality that abstracts from axiological disagreements.

For essentially the same reason, I abstract from disagreements regarding the constituents of ill-desert. Theorists who believe that ill-desert matters—either from the constraint side or because the giving of ill-desert counts as a good—have varied views regarding such matters as whether ill-desert attaches to choices, actions, or character; whether one can have ill-desert by reason of inadvertence; and whether one’s ill-desert is affected by the fortuity that one’s conduct does or does not realize bad consequences. These are questions that philosophers of the criminal law do, and should, continue to debate. But they are most felicitously viewed as debates that arise within a theory of punishment, not as debates over a theory of punishment.

## II. OBJECTIONS

My claim that there exists a near-consensus among contemporary philosophers of criminal law on a single theory of punishment—what I inelegantly term “desert-constrained pluralistic consequentialism” (or “DCPC”)—could be resisted or denied on several grounds. Broadly speaking, however, the varied objections can be usefully grouped into two categories. The first denies that retributivists and consequentialists have converged to the extent I claim. The second contends that, even insofar as consequentialists and retributivists have converged on DCPC, this neat picture of univocality is undermined from the outside so to speak, i.e., by theorists who self-identify as neither retributivist nor consequentialist. The most prominent possible opponents of DCPC, I believe, are proponents of restorative justice.

### *A. Retributivism and Consequentialism Do not Converge*

The claim that contemporary retributivists and consequentialists share the same structure of justificatory argument, and differ only on (likely irresolvable) views about what states of affairs do or do not have intrinsic value, should elicit skepticism. After all, the opposition between retributivism and consequentialism is so deeply seated that some theorists, Feinberg for example, have gone so far as to propose that retributivism be defined simply as a non-consequentialist justification of punishment.<sup>14</sup> Furthermore, as one commentator observed in a different context, “In any important debate, whenever one side declares ‘We are all x now,’ it is a pretty safe bet that the debate has taken a wrong turn—or that someone is trying to pull a fast one.”<sup>15</sup>

---

<sup>14</sup> Joel Feinberg, *What, If Anything, Justifies Legal Punishment? The Classic Debate*, in *PHILOSOPHY OF LAW*, \_\_\_ (Joel Feinberg & Hyman Gross eds. 5<sup>th</sup> ed. 1995).

<sup>15</sup> Gregory Bassham, *Justice Scalia’s Equitable Constitution*, 33 *J.C & U.L.* 143, 154 (2006).

Let us call those putative retributivists who accept the retributive intrinsic good claim (it is intrinsically good that a wrongdoer suffers on account of and in proportion to his blameworthy wrongdoing), and who accept being classed as pluralistic consequentialists, “retributive consequentialists.” Those consequentialists who deny the intrinsic good claim are “non-retributive consequentialists.” (Recall that in emphasizing the convergence on DCPC, my claim to this point is that this distinction within pluralist consequentialism is not especially meaningful or important: We could just as well divide the universe of pluralist consequentialists about punishment in other ways too.) Let us call those retributivists who deny that their view about the justifiability of punishment is accurately captured in consequentialist terms at all “non-consequentialist retributivists.” Non-consequentialist retributivists need not deny (a) that consequentialists about punishment differ on the question of whether the suffering of a wrongdoer is an intrinsic good or (b) that they share something important in common with those who answer that question in the affirmative. Rather, non-consequentialist retributivists insist only that they make use of the claim that wrongdoers deserve to suffer, or to be punished, in a way that is not reducible to any form of pluralistic consequentialism. This Section considers objections to the convergence thesis from retributive consequentialism, non-consequentialist retributivism and non-retributive consequentialism, respectively.

Retributivists might advance several distinct arguments against the claim that the retributive justification for punishment can be fully assimilated to consequentialism. In this section I consider only two, though I am open to the possibility that others should be identified and addressed. The two arguments I consider are: first, that desert-centered consequentialism (as one position within pluralistic consequentialism) is distinct from other forms of consequentialism because the good of deserved suffering is a conceptual or intrinsic consequence of punishment not a contingent one; and second, that the obligation to bring about the good consequences is of a different character for retributivists than for consequentialists. Put another way, the first objection (likely to be mounted by retributive consequentialists) holds that retributivism focuses on a meaningfully different *type of consequence* than do (other) consequentialist theories, and not merely that it

treats different consequences as being of value, or as having greater value, than do its competitors, whereas the second objection (likely to be mounted by non-consequentialist retributivists) holds that retributivism espouses a different *type of relationship* to the consequences upon which it focuses than do consequentialist theses. What would be the non-retributive consequentialist's objections to the convergence thesis are harder to identify. Beyond bemoaning that her felt interest in denying or discrediting retributivism is threatened if retributivism is rendered as just a species of pluralistic consequentialism, the non-retributive consequentialist must maintain that retributivism is committed to one or more claims in addition to the intrinsic good claim that cannot be reconciled with a consequentialist structure of justification.

1. *Of consequences conceptual and contingent.* Several theorists with retributive sympathies who have acknowledged the plausibility of recasting retributivism in consequentialist terms have nonetheless sought to resist what might appear to be the marginalization of retributivism as a distinct punishment theory by proposing that retributive consequentialism invokes a very different type of consequence than does non-retributive consequentialism. George Fletcher pursued this line of argument when distinguishing between "factually" and "conceptually" consequentialist theories.<sup>16</sup> Similarly has Antony Duff distinguished "contingent" from "intrinsic" consequentialism.<sup>17</sup> As Leo Zaibert summarizes:

Not all consequences of a given action are of the same *kind*, and the consequences the retributivist cares about could be seen as different from the consequences which the consequentialist cares about. The consequences that retributivists care about are all *intrinsic* (or inherent, necessary, etc.) to the very act of punishment, whereas the consequences that consequentialists care about are *extrinsic* (or external, contingent, and so on) to the very act of punishment. In other words, both retributivism and consequentialism are consequentialist theories, in a broad sense of "consequence." The way in which retributivism relates to those consequences of punishment that it deems important is a special way: the relation is one of logical necessity. If punishment is justified for the retributivist, it is justified *eo ipso*, necessarily at the very moment in which it is inflicted.<sup>18</sup>

---

<sup>16</sup> George P. Fletcher, *Punishment and Responsibility*, in THE BLACKWELL COMPANION TO THE PHILOSOPHY OF LAW AND LEGAL THEORY \_\_\_, 516 (Dennis Patterson ed. 1996).

<sup>17</sup> R.A. DUFF, PUNISHMENT, COMMUNICATION, AND COMMUNITY (2001).

<sup>18</sup> LEO ZAIBERT, PUNISHMENT AND RETRIBUTION 133 (2006). See also Russell Christopher, *Deterring Retributivism: The Injustice of "Just" Punishment*, 96 Nw. U. L. Rev. 843 (2002).

If retributive consequentialism is intrinsically or conceptually consequentialist and all (or virtually all) forms of non-retributive consequentialism are contingently or factually consequentialist, it would remain to determine how significant the intrinsic/contingent distinction is or, at a minimum, whether foregrounding the distinction illuminates more than it distracts. In fact, though, I am not persuaded that retributivist consequentialism *is* intrinsically consequentialist. To resolve this question requires that we clearly specify just what is this thing—punishment—that we aim to justify and also what is the good state of affairs that retributive consequentialists distinctively value.

We have already addressed the second question when we distinguished between two versions of the desert claim—that wrongdoers deserve to be punished, and that wrongdoers deserve to suffer—and preferred the latter. (When I speak of “deserved punishment,” that will be shorthand for acts of punishment that are intended to, and do in fact, cause deserved suffering.) More particularly, we provisionally endorsed this version of the desert claim: wrongdoers deserve to suffer on account of and in proportion to their blameworthy wrongdoing. The move that paved the way for recasting retributivism as a position within pluralistic consequentialism was to equate this formulation of the desert claim with the claim that it is intrinsically good that wrongdoers suffer on account of and in proportion to their blameworthy wrongdoing.

If we define punishment (roughly) as the inflicting of suffering upon an offender (or supposed offender) for an offense (or supposed offense), then it is true that the intrinsic good of realizing deserved suffering is not a contingent consequence of punishment. It cannot be a contingent consequence of punishment that it cause deserved suffering if punishment is defined as causing deserved suffering. But this is only because we have built the causing of suffering into the definition of the act or practice to be justified. We could instead describe the acts or practices said to stand in need of justification in more fundamental or basic terms (not in the *most* basic terms), as a variety of act-types such as imprisoning or caning. When the retributive consequentialist justifies imprisonment or caning or the like by reference to the suffering that the

act in question is designed to cause, then the consequences invoked are contingent, not conceptual: it is an empirical, not a conceptual or logical, question whether various deprivations of an individual's liberty or various impositions of physical force upon an individual's body in fact cause the individual to endure the negative experience that retributive consequentialists contend is intrinsically good. True, the causal chain is short, internal and exceedingly regular rather than external and subject to substantial empirical contingencies. But these are just matters of degree, not of kind—or, insofar as they are matters of kind, they reflect distinctions whose moral or conceptual significance is obscure.

A retributive consequentialist who wishes to distinguish her own view more sharply from other desert-constrained pluralistic consequentialist approaches might find it significant that the consequences she especially values, even if contingent rather than conceptual, are the most direct or immediate consequences of acts of punishment and, moreover, that the contingent consequences that non-retributive consequences invoke depend upon, or work through, the offender's experience of suffering. However, legal theorists and philosophers are increasingly skeptical of the moral relevance of differences in directness or immediacy of causal chains, and I do not see why anything of importance should be attributed to this claimed distinction. Consider that non-retributive consequentialists who rely heavily on punishment's tendency to reinforce desirable social norms or to bolster general respect for the law probably contemplate longer causal chains between punishment and the intrinsically good states of affairs that justify it than do non-retributive consequentialists who focus on special deterrence. (Few would believe that respect for the law, say, is an intrinsic good; it is instrumentally valuable in helping to bring about other intrinsic goods, often by complex routes.) But I am aware of no theorist who argues that it is meaningful or particularly useful to distinguish consequentialist theories on this basis.

In any event, the premise is not strictly true. Punishment can bring about the good of a reduction in acts of antisocial aggression through the mechanism of general deterrence even if the person punished suffers

not a whit. That others believe that the caning, say, or the imprisonment, if inflicted upon them, would cause them to suffer is sufficient for general deterrence to work its magic. For all these reasons, the distinction between conceptual and contingent consequences seems unpromising as a ground for reviving the centrality of the retributivist/consequentialist divide that is threatened by treating the retributivist's desert claim as equivalent to the intrinsic good claim.

2. *Of justice and intrinsic goodness.* A second retributivist objection to the convergence claim is voiced not by retributive consequentialists but by non-consequentialist retributivists, who would deny that the distinctive feature about retributivism is adequately captured by the intrinsic good claim. Retributivism, after all, is routinely described as a theory of justice: It maintains that justice demands that offenders be given their ill-deserts or, more precisely, that wrongdoers be made to suffer on account of, and in proportion to, their blameworthy wrongdoing. Call this *the retributive justice claim*. It is not equivalent, so the argument must go, to the retributive consequentialist's claim that the state of affairs in which an offender suffers on account of and in proportion to his blameworthy wrongdoing is an intrinsic good that the state has moral reason to bring about.

I find this the most promising objection to the convergence thesis. But it is also the most elusive and therefore requires the most fleshing out. We must start by getting clearer on what it means to claim that justice demands that offenders be given their ill-desert. As the earlier passage from Moore suggests, the retributive justice claim is often cast in terms of society's duty or obligation to bring about deserved suffering. The contention that we have an *absolute* obligation to punish offenders, or to seek to realize deserved suffering, is often attributed to Kant. But regardless of whether Kant did in fact espouse such a view—and the frequently voiced contention that he did is much disputed—I am aware of no contemporary philosopher or criminal law theorist who advances such an extreme position. Accordingly, a far more plausible construal of the retributive justice claim holds that we have a *pro tanto* obligation to bring about deserved suffering, one

that can be defeated only by certain sorts of considerations. As a pluralist, the retributive consequentialist will recognize various types of intrinsic goods and bads, many of which have the potential, singly or in combination, to direct that the state ought not to punish notwithstanding the good that such punishment would realize in the coin of deserved suffering. If the retributive justice claim cannot be captured by a suitably nuanced formulation of the retributive intrinsic good claim, it must be, I think, because some types of reasons, no matter their weight, are simply of the wrong sort to enter into the moral calculus as reasons against the doing of justice. In the Razian vocabulary, the non-consequentialist retributivist is likely to claim that seeing that justice is done—in this case, that *retributive* justice is done—is a “protected reason” to punish (supposed) wrongdoers, which is to say that it is a consideration that both serves as a first-order reason to punish (in fact, a first-order reason of substantial weight) and has the second-order function of excluding from consideration some (but not all) of what would otherwise be sound first-order reasons not to punish.

This objection seems highly plausible in principle, by which I mean not that the retributive justice claim is highly plausible (about which I am agnostic for present purposes), but only that it is highly plausible that the retributive justice claim can be functionally nonequivalent to the retributive intrinsic good claim. In evaluating whether the two claims are likely to be functionally nonequivalent in practice, my strategy will be to catalogue the principal types of considerations that a pluralistic consequentialist might recognize as sound reasons not to impose deserved suffering in a given case, and then to identify which if any of those reasons are plausible candidates to be *excluded* as reasons against our pro tanto obligation to do justice.

As best I can tell, five sorts of reasons are most commonly relied upon as arguments against criminal punishment, either globally or in particular cases:<sup>19</sup> (1) that an offender has a right against the infliction of suffering, or the intentional deprivation of certain important liberty interests, even assuming *arguendo* that he deserves to experience such suffering or (what we have assumed to be equivalent) that for him to experience

---

<sup>19</sup> See generally Husak, *Why Punish the Deserving?*; Honderich.

such suffering or deprivation would be intrinsically good; (2) that forgoing the infliction of deserved punishment on one offender might be necessary to enable the infliction of deserved suffering on others, as in the case when the state needs the testimony of one offender to secure evidence required to convict another; (3) that imposing what is believed to be deserved punishment always risks imposing suffering either on one who lacks ill-desert entirely or in excess of what the wrongdoer deserves; (4) that inflicting deserved punishment on a wrongdoer often causes undeserved suffering to discrete identifiable innocent individuals who love, or depend upon, the wrongdoer; and (5) that establishing and operating a system of punishment is expensive, consuming material resources that could be put to other social ends, including being left in the hands of taxpayers. Our question is whether any of these considerations is both (a) cognizable by retributive consequentialists as a reason not to punish that could (alone or in combination) *potentially* defeat the reason to punish supplied by the intrinsic good claim, and (b) not similarly cognizable by non-consequentialist retributivists wielding the retributive justice claim. My hypothesis is that the non-consequentialist retributivist embrace of the retributive justice claim threatens the convergence thesis only insofar as we can identify reasons against punishment that satisfy both criteria.<sup>20</sup>

The first consideration against punishment is likely to have force for neither retributive consequentialists nor non-consequentialist retributivists. Surely non-consequentialist retributivists will deny that punishment violates an offender's rights. But, for reasons I have elaborated on elsewhere,<sup>21</sup> retributive consequentialists are able to reach that conclusion as well. Very briefly, I believe that the customary panoply of moral rights (the right to bodily integrity, the right against deprivations of liberty, etc.) derive from the single foundational right to be treated with respect as a person, where respect is a quality of a person's attitudes, or attitudes-in-action, toward others. Unfortunately (or perhaps fortunately), what conduct an appropriate attitude of respect requires of us in a given case is often indeterminate. People have preferences

---

<sup>20</sup> [Dear Reader: I suspect that much in this section—and, possibly, much in the entire paper—depends upon this hypothesis. If you believe it mistaken, I'd be especially grateful for your views.]

<sup>21</sup> See Mitchell N. Berman, *Punishment and Justification*, 118 *Ethics* 258, 278-84 (2008).

and interests and are also choosing beings. But because personhood is not reducible to preferences, interests, capacity for agency, or anything else, an individual's right to respect as a person similarly does not translate into a right that regard be paid *especially* to any one of these aspects or dimensions of personhood as opposed to another. Because wrongdoers experience suffering as a bad, it is likely that one way of respecting them is to refrain from causing them pain. But insofar as they have exercised their wills to violate legitimate interests of others, it is also plausible that causing them to suffer on account of their willing also respects them. Here as is so often the case (but so often overlooked), our duties depend upon our motivations. We respect a person's nature as sentient being by refraining from causing her pain. We respect a person's nature as responsible agent by giving her what she deserves by virtue of the exercise of her will. If we are properly motivated, we can do either—not punish because it is painful, or punish because it is deserved—and still accord her the respect that she is due. If this is correct, then to punish somebody with the goal of inflicting deserved suffering does not infringe that person's right to respect as a person if the punishment issues from reasonable regard for that person's autonomous exercise of her capacity for responsible choice.

Just the opposite seems true of the second through fourth considerations: each should be recognized as a sound reason (of some weight) not to punish for proponents of the intrinsic good claim and the retributive justice claim alike. This is most plain, perhaps, in the case of the second consideration: while retributive consequentialists will trade off between the good of imposing deserved suffering upon A and B, non-consequentialist retributivists will see the trade off as between doing justice to A and doing justice to B. That non-consequentialist retributivism might be described as deontological or agent-relative is no bar, for in the case of both parties, A and B, it is the same agent who will or will not do justice.

Much the same, I believe, is true of the third consideration. Virtually all consequentialists whose theory of value is not exclusively welfarist will view the infliction of punishment on the innocent, and the infliction of punishment in excess of an offender's desert, as especially grave bads—i.e., as bads far worse than

obtain when an individual experiences the same amount of suffering by natural causes. Of course, to accept the desert constraint is to accept that to knowingly inflict punishment on the innocent is forbidden—either absolutely, or in all but the most exceptional (hypothetical) circumstances.<sup>22</sup> But a broad consensus holds that this is a bad much to be avoided even when not done knowingly. Retributive consequentialists should have no principled objection trading off the bad of undeserved suffering against the good of deserved suffering. If they view the bad of undeserved suffering as greater than the good of deserved suffering, they will urge that the various doctrines and practices of the criminal law be structured to substantially reduce the probability of undeserved punishment, as, for example, by endorsing a heightened standard of proof and by resisting strict liability offenses. I am aware of no non-consequentialist retributivist who does not make a similar calculation. None, for example, advocates that the burden of proof be reversed, such that juries be instructed to convict unless persuaded beyond a reasonable doubt of the defendant’s factual innocence, or that offenses be redrawn to eliminate culpability. This suggests that the non-consequentialist retributivist recognizes, as a principle of justice (or as a moral principle of a character admissible against arguments of justice) that the state take substantial pains not to punish the innocent.

But if that is so, then it seems unlikely that the retributive consequentialist and the non-consequentialist retributivist need differ on whether the fact that punishment would cause foreseeable harm to innocent third parties may count as a reason against the infliction of deserved punishment.<sup>23</sup> We have just supposed that the fact that we might be wrong about a defendant’s desert constitutes a reason (of some weight) for both retributive consequentialists and non-consequentialist retributivists not to impose what they believe would be deserved punishment. But the consideration presently under discussion similarly gains its force from our supposed obligations not to cause undeserved suffering. There are two salient differences between the two cases, which might seem to cut in different directions. First, the state causes the innocent defendant to suffer intentionally, but causes the innocent third party to suffer only knowingly, not

---

<sup>22</sup> See Moore on “threshold deontology”

<sup>23</sup> See Jeffrey Brand-Ballard.

intentionally. Proponents of the doctrine of double effect would treat this as a morally significant difference. Second, while intentionally causing the innocent defendant to suffer, the state lacks knowledge about the feature—the defendant’s innocence—that renders its intentional causing of suffering a grievous bad or wrong. In contrast, when knowingly (albeit indirectly) causing suffering to innocent third parties, the state will sometimes know or believe that the third party’s suffering is undeserved. There is a sense in which these two formally distinct aggravating factors—the intentional (as opposed to mere knowing) causing of harm, and the knowledge (as opposed to mere awareness of risk) that the harm caused is undeserved—are made of similar stuff, or have a similar texture. Of course, more needs to be said on this score. Provisionally, though, it is hard to imagine the plausible account of political morality that would consider the suffering experienced by innocent identifiable third parties wholly inert as a reason not to bring about the justice of retributive punishment in a given case.

Let us briefly take stock. The retributive consequentialist necessarily treats the bringing about of deserved suffering as a weighty good that helps to justify the infliction of criminal punishment in a given case. As a pluralist consequentialist, she also recognizes other goods that help to justify criminal punishment. Moreover, and again as a pluralist, she will recognize that not punishing either brings about goods or (what is analytically the same) avoids bads that might make it the case that, all things considered, the state ought not to punish in a given case. These three reasons—to enable the state to impose deserved suffering on others, to reduce the chance of imposing undeserved punishment, and to reduce the infliction of undeserved suffering on third parties—all seem to be sound reasons against punishment, in a given case, even for the non-consequentialist retributivist who would justify punishment only in virtue of the state’s pro tanto obligation to see that retributive justice be served. Up to this point, then, the effort of the non-consequentialist retributivist to distance herself from DCPC does not seem compelling.

The fifth proposed reason against punishment seems more promising as the sort of reason that weighs in the calculus of the retributive consequentialist but not of the non-consequentialist retributivist. The criminal justice apparatus is hugely expensive. Even assuming *arguendo* that deserved punishment promotes the good of deserved suffering, the resources it consumes in doing so could be put to a great many other uses that would serve a diversity of goals, including such quotidian ends as increasing welfare and satisfying preferences. As a value pluralist, the retributive consequentialist is bound to recognize promotion of interests such as these as potential reasons to eschew or reduce deserved punishment (unless she were to deny, implausibly, that welfare and utility have intrinsic value at all). Things are different for the non-consequentialist retributivist. If reasons of justice can be opposed only by some subset of what are generally sound moral reasons—if, as I am assuming, that is the point or function of a moral reason being a reason of *justice*—then the reasons most plausibly excluded are reasons of ordinary welfare and preference-satisfaction, and the like. In short, that punishment is costly (in the ordinary sense) is a reason not to punish for the retributive consequentialist but not for the non-consequentialist retributivist.

I believe that conclusion is too quick, even assuming that ordinary welfare gains do not supply reasons of any weight against punishment for the non-consequentialist retributivist. Given the varieties of justice and the fungibility of money, there is no reason why the opportunity costs that criminal punishment incurs need be cashed out in welfarist terms. Even granting, with the non-consequentialist retributivist, that the infliction of deserved suffering (under appropriate circumstances) serves or constitutes a form of justice, no contemporary theorist is likely to deem that the *only* form of justice; distributive justice, corrective justice, and social justice, perhaps among others, all count too. But the pursuit of retributive justice can come at the expense of these other forms of justice in two ways. First, by decreasing an offender's welfare, the actual infliction of punishment directly implicates concerns of distributive justice in addition to concerns of retributive justice. Or

at least so it is plausible to think.<sup>24</sup> Second and more to the present point, the resources that a state uses to produce retributive justice could be employed instead to promote other forms of justice. Put simply, because the financial cost of punishment can be cashed out not only in welfarist terms, but also in justice-related terms, that criminal punishment is so expensive furnishes a reason against the pursuit of retributive justice, even for the non-consequentialist retributivist, in any state that could use those resources to better promote other forms of justice.

If this analysis is correct, then it does not yet appear that non-consequentialist retributivism excludes from the punishment calculus any considerations that retributive consequentialism recognizes. True, non-consequentialist retributivists might give little weight to some of these potentially competing considerations. But so too might retributive consequentialists; both structures of reasoning can accommodate the full range of possible judgments regarding the weight or force of a consideration that it recognizes. If the retributive desert claim and the intrinsic good claim are functionally equivalent, we have not yet seen adequate grounds to conclude that they are not equivalent as well (at least under circumstances of nonideal theory in which other forms of justice remain imperfectly realized) to the retributive justice claim.

3. *On the logical distinctiveness of retributivism.* Let us finally consider an objection to the convergence thesis from the perspective of non-retributive consequentialism. Recall Moore's assertion that the intrinsic good claim captures "what is distinctively retributivist."<sup>25</sup> "At first glance," Dolinko objects, "one might suppose that Moore's assertion is patently false, because a non-retributivist might endorse the intrinsic good claim."<sup>26</sup> Accordingly, he contends, retributivism must have some sort of distinctive logical structure; it is not enough that it commit to a particular view of intrinsic goodness:

---

<sup>24</sup>For discussion, see Douglas Husak, *Holistic Retributivism*, 88 Cal. L. Rev. 991 (2000); Richard J. Arneson, *Desert and Equality*, in *EGALITARIANISM* (N. Holtug & K. Lippert-Rasmussen, eds. 2006); Saul Smilansky, *Control, Desert and the Difference Between Distributive and Retributive Justice*, 131 Phil. Stud. 511 (2006).

<sup>25</sup> See supra tan \_\_\_\_.

<sup>26</sup> Dolinko, *The Intrinsic Goodness of Punishment*, at 517.

Rather, what is distinctive about the retributivist must be the *role* played in her theory by the intrinsic goodness of punishing the guilty. For the retributivist, this intrinsic goodness cannot be an irrelevancy or a mere happy accident. It must be either (i) the reason for engaging in the practice of punishment (its rational justification), or (ii) the reason why that practice is morally permissible (its moral justification), or both.<sup>27</sup>

Now, it is not entirely clear why the retributivist *must* assign to the intrinsic goodness of punishing the guilty (or, as I would put it, the intrinsic goodness of a wrongdoer's suffering on account of and in proportion to his blameworthy wrongdoing) a distinct role in her justificatory account. It could just be that the most plausible form of retributivism, and the version most widely embraced today, is a species of consequentialism. It could be, in other words, that self-described contemporary retributivists *just are* those consequentialists about punishment who believe that deserved suffering has sufficient intrinsic value to significantly affect our thinking about how to constitute the doctrines and practices of the criminal justice system.<sup>28</sup> The convergence thesis suggests that this is precisely the case. Nonetheless, the retributive intrinsic good claim might still play a role in the justificatory logic of punishment different from the role played by other intrinsic goods. So while I will argue that a retributivist need not believe that giving deserved suffering constitutes *the* (sole) justification (rational or moral) for punishment in the way that Dolinko has in mind, he is right that there is something special about the role that the retributive intrinsic good claim plays in the justificatory enterprise.

To understand the argument, we must first distinguish two different types of justification, what I will call "all-things-considered" and "tailored."<sup>29</sup> An all-things-considered justification of an act or a practice is just what it sounds like: it establishes that the act or practice is morally justified, or permissible, in light of all considerations. A tailored justification, in contrast, establishes the permissibility of the challenged act or practice against one or more particular grounds for doubt, what I have termed "demand bases."

---

<sup>27</sup> Dolinko, *The Intrinsic Goodness of Punishment*, at 518.

<sup>28</sup> Possibly, however, retributive consequentialists are wrong about that. Consequentialists who deny the intrinsic good claim and even the desert constraint, but who appreciate the instrumental value of structuring the criminal justice system to accord with popular beliefs in the importance of giving deserved punishment, could end up backing doctrines and policies that closely approximate what a retributive consequentialist might advocate. See *generally* Robinson & Darley, *The Utility of Desert*.

<sup>29</sup> The analysis to follow draws from Berman, *Punishment and Justification*.

As our analysis of the non-consequentialist retributivist's justice-based objection to the convergence thesis suggests,<sup>30</sup> the retributive consequentialist must rely on a plurality of values insofar as she aims to justify punishment all things considered. Historically, however, the core concern of philosophers of punishment has *not* been to provide an all-things-considered justification. Rather, it has been to meet the very particular reason why punishment has been understood to stand in special need of moral justification. As Hart put it, agreeing with Benn, it is "the deliberate imposition of suffering which is the feature needing justification."<sup>31</sup> That is, theorists routinely observe not merely that "punishment stands in need of justification"—which contention might be an invitation for all-things-considered justification—but that such justification is needed precisely "because it involves the infliction of pain or other form of unpleasant treatment."<sup>32</sup> As another commentator rightly emphasized, "The moral problem that the having of a legal institution of punishment presents can be stated in one sentence: It involves the deliberate and intentional infliction of suffering."<sup>33</sup> Because it is the fact that punishment involves the intentional infliction of suffering that *particularly* demands justification, philosophers of punishment have understood their core task to be explaining how punishment can be justified in light of, or against, *this* objection. They have, in other words, sought a justification of punishment tailored to meet this concern, even if other (subsidiary) objections to punishment—e.g., that it is costly—might remain to be addressed before the all-things-considered justifiability of punishment can be established.

---

<sup>30</sup> See *supra* Section II.A.2.

<sup>31</sup> HART, PUNISHMENT AND RESPONSIBILITY, p. 2 n.3.

<sup>32</sup> Claire Finkelstein, *Positivism and the Notion of an Offense*, 88 Cal. L. Rev. 335, 358 (2000). A tiny sampling of similar claims includes IGOR PRIMORATZ, JUSTIFYING LEGAL PUNISHMENT 7 (1989) ("To punish means to inflict an evil. But to inflict evil on someone is something that, at least *prima facie*, ought not to be done. So the question arises: What is the *moral justification* of inflicting the evil of punishment on people? . . . This is the question about punishment which is being discussed in philosophy . . ."); Richard Wasserstrom, "Why Punish the Guilty?" in *Princeton University Magazine* 20 (1964): 14-19, reprinted in *Philosophical Perspectives on Punishment*, ed. Gertrude Ezorsky (1972): 328-41, p. 337 ("Punishment is an evil, an unpleasantness; it requires that someone suffer. Its infliction demands justification."); NICOLA LACEY, STATE PUNISHMENT: POLITICAL PRINCIPLES AND COMMUNITY VALUES 13 (1988) ("The most obvious reason for a need to justify punishment is that it involves, on almost any view of morality, *prima facie* moral wrongs: inflicting unpleasant consequences . . . and doing so irrespective of the will or consent of the person being punished.").

<sup>33</sup> R.W. Burgh, *Do the Guilty Deserve Punishment?*, 79 J. Phil. 193, 193 (1982).

Seeking an all-things-considered justification for punishment,<sup>34</sup> Dolinko concludes that the retributive intrinsic good claim is not up to the task and, indeed, plays no special or distinctive role relative to the other types of intrinsic goods that a pluralistic consequentialist is likely to invoke. I believe Dolinko is right about this. But he is wrong, I believe, when we shift our focus to the search for a tailored justification. Very simply, against the demand that punishment be justified because it inflicts pain or suffering on wrongdoers, the intrinsic good claim justifies punishment by cancellation, whereas other good consequences that the pluralist might call upon proceed by override. That is, if the suffering of a blameworthy wrongdoer is an intrinsic good, then the principal mark against punishment simply dissolves; it lacks moral force. But if the suffering of a blameworthy wrongdoer is intrinsically bad—as one who denies the retributive intrinsic good claim will maintain—then the consequentialist reasons to punish can supply even a tailored justification of punishment only by outweighing the principal standing reason not to punish. Because cancellation enjoys logical priority to override in an argumentative dialectic, there *is* something distinctive about the role played by the retributive intrinsic good claim that is not played by other intrinsic goods, which, if not exactly what Dolinko demanded, is close enough.

Put another way, once we recognize the centrality, in the philosophical literature, of the search for tailored justifications for punishment, then we have at least some reason to agree with Moore that “what is distinctively retributivist is the view that the guilty receiving their just deserts is an intrinsic good.” In this way, there remains something importantly (or at least interestingly) distinct about retributivist and consequentialist accounts of the justifiability of punishment—even if, as per the convergence thesis, the partition between retributivists and consequentialists is now best drawn between retributive consequentialists and non-retributive consequentialists.

*B. Desert-Constrained Pluralistic Consequentialism is Denied by Restorative Justice*

---

<sup>34</sup> See, e.g., Dolinko, *The Intrinsic Goodness of Punishment*, at 521 (inquiring whether retributivism can “rationally justify the practice of punishment in the fact of its staggering costs”).

The previous section considered and found wanting three arguments to the effect that retributivists and consequentialists have not converged, to the degree I claim, on a theory or justification for punishment. This section address the very different argument that, even to the extent that retributivists and consequentialists have converged on DCPC, there remain other theories of punishment—neither retributivist nor consequentialist—that have not. I suppose that, in principle, there could be any number of external competitors to DCPC. But I believe that the “restorative justice” movement constitutes the only prominent contemporary competing view. So the issue to be considered here is simply whether the existence of restorative justice theories or approaches to punishment significantly undermines the convergence thesis.

The short answer is that it does not. And the reason is that restorative justice is not a true competitor even to classical retributive or consequentialist theories of punishment because it addresses itself to a different question than do they. The question that retributivists and consequentialists have endeavored to address is this: what, if anything, justifies the state in subjecting individuals to criminal punishment given (as we have just emphasized) that it involves the intentional infliction of suffering? Restorative justice is a framework for an answer to this rather different question: what should the state do with offenders, or in the face of criminality? The answer it provides is that, in many cases, the state should *not* punish the offender but should instead seek to repair relationships that have been damaged and to meet legitimate needs of victims and wrongdoers, paradigmatically (though not exclusively) through victim-offender mediation.

As we have seen, the pluralistic consequentialist answer to the question that it addresses is that punishment is justified because and insofar as it produces net good consequences—however the goodness of consequences may be measured by one’s particular moral theory, or theory of value. Insofar as punishment produces less good consequences than would an alternative response to criminal wrongdoing, it necessarily follows that the state should not inflict punishment. To be sure, many pluralistic consequentialists, including retributive consequentialists, might be less sanguine than are proponents of restorative justice regarding how

well non-punitive measures are likely to advance social interests that count in their joint calculus. But resolution of any such disagreements depends more upon highly context-specific evaluative and predictive judgments than upon philosophical disputation. The contrary conclusion instantiates a phenomenon in the philosophy of punishment remarked upon nearly fifty years ago—namely, the “great deal of confusion and many false oppositions between theories which are not simply different but which are attempting to solve different problems.”<sup>35</sup> In sum, then, there is nothing about restorative justice, rightly understood, that should require its proponents to disagree with desert-constrained pluralistic consequentialists regarding what *would* justify punishment. What drives proponents of restorative justice is the conviction that, given the ability of properly designed non-punitive alternatives to achieve shared social ends, punishment is in fact justified less often than it is practiced.

### III. CHALLENGES

Part II identified and evaluated objections to the claim that contemporary criminal law theory has nearly converged on a view about the justification for criminal punishment that I have styled desert-constrained pluralistic consequentialism. The objections were, first, that at least one robust form of retributivism is not faithfully rendered as any form of consequentialism and second, that even insofar as retributivism and consequentialism have converged in this way, they are now supplemented by a third theory of punishment (marching under the banner of “restorative justice”) that does not. This Part shifts attention from the convergence thesis to DCPC itself. Even if DCPC does not enjoy quite as much contemporary support as I have claimed for it, few if any readers are likely to doubt that it is accepted by a great many contemporary theorists. This Part draws attention to three of the most underappreciated yet difficult challenges it faces—

---

<sup>35</sup> K.G. Armstrong, *The Retributivist Hits Back*, 70 *Mind* 471, \_\_\_ (1961).

challenges that arise from the theory's concern with ill-desert, either as a possible good to bring about, or as a necessary condition that constrains the exercise of criminal punishment.<sup>36</sup>

Let me make clearer what sort of challenges I have in mind by indicating, if loosely, what I don't. First, I am not presently concerned with the myriad debates regarding how to conceptualize or operationalize ill-desert that, while they must be resolved, can be resolved in any of a number of ways none of which would significantly threaten the viability or plausibility of DCPC. In this category would lie such questions as whether one can have ill-desert for negligence or whether the fortuitous realization or non-realization of intended or contemplated harms can affect the magnitude of one's ill-desert. Although different answers to questions such as these do entail different doctrinal recommendations, none of the answers presently on offer seems likely to undermine the coherence or attractiveness of a desert-constrained or desert-inclusive consequentialism. The determinist challenge to moral responsibility presents a stark contrast. If the challenge cannot be satisfactorily met on libertarian or compatibilist grounds, then DCPC's emphasis on desert would seem untenable. This, then, is the sort of challenge I have in mind—or (and here's my second ground of distinction) it's what I would have in mind except for the fact that there is no dearth of philosophical attention to the subject. Nobody needs this essay to suggest that a solution to the global puzzle of moral responsibility would be welcome.

That said, here are three issues that have the potential to present genuine and significant challenges to our dominant justification for criminal punishment, but that, if not entirely overlooked, have not attracted the degree of attention that, it seems to me, they warrant.

*1. Already suffered enough?* How, if at all, does an offender's nonpenal suffering affect the amount of suffering he deserves on account of his blameworthy wrongdoing? Distinguish two ways in which it might.

The first is what Gertrude Ezorsky dubbed the "whole life" view of suffering: how much suffering a wrongdoer

---

<sup>36</sup> [Dear Reader (again): Although this Part identifies only 3 challenges at present, I expect to add a few more as they occur to me—or (more to the point) as they may be suggested to me.]

deserves must be a function of his life-long ledger of pleasures and pains. This view can also be understood, perhaps, as an effort to effectuate a closer integration between retributive and distributive justice.

A second and (to my eyes) more plausible view maintains that, even if penal suffering ought not to be adjusted for pre-offense suffering, it should be reduced in some fashion to account for suffering that one experiences as a result of his blameworthy wrongdoing.<sup>37</sup> Consider, for example, Michael Vick, who lost somewhere in excess of \$100 million in lost salary and endorsements, not to mention fame and glory, as a consequence of his dog-fighting. May a desert-sensitive criminal justice system disregard the non-penal suffering he has already endured on account of his blameworthy wrongdoing when selecting a sentence? Or consider this chestnut: A is wrongly charged and convicted for a crime he did not commit. After 10 years imprisonment, he is exonerated and released. He later commits an offense. Does his previous time served count, in any way, against what he deserves to suffer now?

*2. Knowingly punishing in excess of desert.* Theorists often equate punishment in excess of ill-desert with punishment absent ill-desert, contending that after an offender receives the suffering he deserves any additional punishment is functionally and morally identical to punishing someone who lacks ill-desert entirely. The claim has a superficial plausibility, but is far from self-evident. Query: is a 10-year sentence imposed on A, whose ill-desert warrants (we believe) a sentence only half as long fully morally equivalent to a 5-year sentence imposed on B, who (we believe) is wholly innocent? Simply put, the relationship between punishing offenders more than they deserve and punishing persons who deserve no punishment at all warrants more careful scrutiny than, as far as I am aware, it has received.

*3. Risking punishing absent ill-desert.* The desert constraint maintains that it is impermissible to knowingly punish someone who lacks ill-desert (for the offense for which he is charged) and also (subject to

---

<sup>37</sup> This “as a result” formulation is imprecise. The suggestion is not, I believe, that a bank robber who breaks his leg during flight should be punished less on account of his injury. Rather, the wrongdoer’s suffering must come about in the right sort of way—caused, we might say, by the wrongfulness of his action.

doubts just raised) to knowingly punish someone more than his ill-desert warrants. As discussed earlier,<sup>38</sup> the constraint does not bar the infliction of punishment under circumstances in which the punisher is aware of the mere possibility that the suffering to be inflicted exceeds ill-desert (including because the putative offender has no ill-desert at all). Yet even if the desert constraint does not itself apply, surely kindred moral norms substantially limit the permissibility of *risking* punishing in excess of desert. What are the nature and content of any such proscriptions? Are they the very same considerations that are thought to support the “beyond a reasonable doubt standard” of proof? And if they are, why would the magnitude of (epistemic) risk that marks the threshold between permissible and impermissible risk-taking be invariant across contexts?

#### CONCLUSION

[To come]

---

<sup>38</sup> See argument (3) in Section II.A.2.