University of Pennsylvania Carey Law School
Penn Law: Legal Scholarship Repository

Faculty Scholarship at Penn Law

2011

Two Kinds of Retributivism

Mitchell N. Berman
University of Pennsylvania Carey Law School

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship

 Part of the Criminal Law Commons, Ethics and Political Philosophy Commons, Law and Philosophy Commons, Law Enforcement and Corrections Commons, and the Public Law and Legal Theory Commons

Repository Citation

Berman, Mitchell N., "Two Kinds of Retributivism" (2011). *Faculty Scholarship at Penn Law*. 2353.
https://scholarship.law.upenn.edu/faculty_scholarship/2353

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.

TWO KINDS OF RETRIBUTIVISM

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.

This chapter aims to assess the state of that debate in the early years of the 21st century. It advances one principal claim. Additionally, and consistent with this volume's ambition to help set terms

* Richard Dale Endowed Chair in Law, Professor of Philosophy (by courtesy), the University of Texas at Austin.
Earlier drafts of this paper were presented at a GALA Workshop at Berkeley Law; at the conference for this book, held at Rutgers-Newark School of Law; and at UT Law's law and philosophy discussion group. I am indebted to participants at those events and wish especially to recognize valuable criticisms and suggestions from Larry Alexander, Mike Cahill, Meir Dan-Cohen, John Deigh, Antony Duff, David Enoch, Kim Ferzan, Stephen Galoob, Sandy Kadish, Chris Kutz, David Sklansky, Sarah Song, Victor Tadros, Kevin Toh, and Peter Westen. Guha Krishnamurthi provided excellent research assistance. With apologies to Michael Smith, 'Two Kinds of Consequentialism' (2009) 19 Philosophical Issues 257.

and direction for fruitful philosophizing about the criminal law in the near future, it proposes an agenda for further work.

The central argument of this paper is that the dominant classificatory framework of theories of punishment that had become orthodox by the latter decades of the 20th century is in peril. That taxonomy was centered on a two-part distinction between consequentialist and retributivist justifications for punishment. To state very precisely what either of these approaches maintains would beg a great many debates. As a first pass, though, consequentialist theories (what I will later rename instrumentalist theories) justify punishment by the good that punishment produces, whereas retributivist theories see punishment as justified by reference to the wrongdoers' supposed ill-desert—the claimed fact that wrongdoers deserve to suffer, or to be punished, or something of this sort. I argue that this framework is no longer accurate because retributivism has increasingly morphed into an account that rests upon a justificatory structure that is plainly consequentialist. That is, it seems increasingly fitting to view retributivism as a subtype of consequentialist justifications for punishment—a "retributivist consequentialism" that can be meaningfully contrasted with varieties of "non-retributivist consequentialism"—rather than as an alternative to them. Some will think this a contradiction in terms, that retributivism is non-consequentialist by definition. On this view, "retributivist consequentialism" is oxymoronic. I will argue that that is not so.

That there exists a theory or account of the justifiability of criminal punishment that is both recognizably retributivist and assimilable to a consequentialist structure of justification does not mean that that account exhausts retributivist space. Even if, as I argue, it is meaningful to speak of retributivist consequentialism, there might nonetheless exist other tenable forms of retributivism that resist reduction to consequentialism. If so, we should recognize two kinds of retributivism, not one: consequentialist retributivism (which is simply an alternate name for what I have just referred to as retributivist consequentialism) and non-consequentialist retributivism. At this moment in time, though,

the consequentialist variant of retributivism is more perspicuous.[1] Accordingly, among the more pressing tasks for philosophers of the criminal law who continue to be interested in what has long been the central issue in the field—what justifies the infliction of criminal punishment?—is to explicate non-consequentialist retributivism in its most persuasive and attractive light—if not necessarily to establish that it *is* true, then at least to show how it could be.

1. 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 at least appears to mirror 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

[1] For a valuable recent discussion of "consequentialist retributivism"—one that observes that, despite its "intuitive appeal . . . , it has thus far apparently received *no* explicit, sustained defense in the scholarly literature," and explicates some of its advantages as against standard (non-consequentialist) retributivism, see MT Cahill, 'Retributive Justice in the Real World' (2007) 85 Wash U LR 815, 825, 833-40.

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.[2]

Now, the facile mapping of the retributivist/consequentialist divide in punishment theory onto the deontological/consequentialist divide in moral theory is highly deceptive. Even if (as I am willing to suppose, but only arguendo) retributivists about punishment cannot be consequentialists about morality generally, the converse is clearly false: consequentialists about punishment need not endorse a consequentialist comprehensive moral theory. That is, they need not entirely eschew deontological or non-consequentialist aspects either of normative ethics or of political morality. A justification of punishment is conventionally labeled "consequentialist" if it explains the moral permissibility of the practice by reference to the value of its actual or supposed consequences. But there is no reason why the consequences that are thought to matter must themselves be conceptualized in resolutely consequentialist terms. For example, one can justify punishment on consequentialist grounds while still arguing that it is more important to deter intentional killings than mere allowings-to-die on the grounds that the former are deontological moral wrongs and the latter aren't. Furthermore, many persons who ascribe to what are traditionally termed consequentialist theories of punishment grant that the state is nonetheless under an obligation of political justice not to punish persons who aren't blameworthy, and they might believe as well (though they need not) that constraints of this sort cannot be

[2] This two-part division is so dominant that supporting citations are surely unnecessary. A small and close to random sampling of classificatory schemata based on an opposition between consequentialism (sometimes less helpfully labeled "utilitarianism") includes RA Duff, *Punishment, Communication, and Community* (2001) ch. 1; I Primoratz, *Justifying Legal Punishment* (1989); CL Ten, *Crime, Guilt, and Punishment* (1987); K Greenawalt, 'Punishment' in SH Kadish (ed) 4 *Encyclopedia of Crime and Justice* 1336 (1983).

consequentialized.[3] For reasons such as these, consequentialism about punishment is not identical to consequentialism about morality.

Unfortunately, there is no obviously best solution to this problem: while it would have been preferable had philosophers about punishment not drawn their classificatory line between retributivism and "consequentialism" these many years (so as to forestall the mistaken assumption that consequentialists about punishment must be consequentialists about morality generally), to change nomenclature now risks a different confusion (that the new term refers to something different from what "consequentialism" had previously been thought to mean in this context). On balance, though, the latter risk seems worth running. Following Victor Tadros,[4] then, I will apply the term "instrumentalist" to the forward-looking justifications of punishment that tradition contrasts with the backward-looking justifications labeled "retributivism." That is, let us for the moment accept whole hog the familiar distinction between retributivist and consequentialist justifications for punishment while renaming the latter "instrumentalism." I spoke in the Introduction of retributivist and non-retributivist consequentialism, and of consequentialist and non-consequentialist retributivism. Those very same ideas we can now re-caption as retributivist and non-retributivist *instrumentalism*, and as *instrumentalist* and *non-instrumentalist* retributivism, without any intended change in meaning. That is the nomenclature I will employ for the remainder of this chapter.

2. Retributivist Instrumentalism

2.1 Retributivism and Desert

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

[3] On efforts to consequentialize constraints on the pursuit of consequentialist objectives see below n 40.
[4] V Tadros, *The Moral Foundations of Criminal Law* (forthcoming). For reasons that would require an extended digression to explicate, I am not persuaded that Tadros and I use the term "instrumentalism" to mean precisely the same thing. But any such difference between our usages as might exist is a matter of detail.

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.[5] 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.

Over the ensuing years, however, a consensus has arisen. As CL Ten put it in a much-read book, "Contemporary retributivists treat the notion of desert as central to the retributivist theory, punishment being justified in terms of the desert of the offender."[6] Or, in Larry Alexander's words: "Retributivists argue that punishment must be justified by the ill-desert of the one punished."[7] Let us call this core retributivist contention *the desert claim:* punishment is justified by the offender's ill-desert.

Desert talk is notoriously mysterious and elusive. It provokes many questions. Here are two. First, just what is that offenders deserve? Second, what does it mean that he deserves it?

Joel Feinberg was the first to analyze desert as a triadic relationship: an agent, that which is deserved, and that which makes the agent deserve it.[8] He called the last the "desert basis" but didn't offer pithy labels for the other two.[9] Let's call them the desert subject and the desert object, respectively. Our first question, then, can be put this way: What—precisely—is the retributivists' desert object?

[5] J Cottingham, 'Varieties of Retribution' (1979) 29 Phil. Q. 238.
[6] Ten (n 2 above) 46.
[7] L Alexander, 'The Doomsday Machine: Proportionality, Punishment and Prevention' (1980) 63 Monist 199, 199.
[8] J Feinberg, *Doing and Deserving: Essays in the Theory of Responsibility* (1970).
[9] But see n 53 below.

This question is provoked by the marked vagueness of *the desert claim.* To say (as the passages quoted above do) that desert is "central" to retributivism, that punishment is justified "in terms of" or "by" the offender's desert, does not specify just what it is that the offender deserves. Yet more significantly, when the desert object *is* specified, it is articulated in differing terms. Time and again, it is said that, for retributivists, "punishment is justified because people deserve it."[10] The antecedent of "it" being punishment, this formulation implies that the retributivist desert object is punishment itself. Other scholars, however, declare that "all retributive theories assert that offenders deserve to suffer."[11] On this view, the retributivist desert object is *suffering,* not *punishment.* That differing descriptions of the retributivist desert claim can be found need not imply that the issue is a subject of genuine debate. To the contrary, I submit that most readers of the literature would conclude that commentators have rarely given explicit attention to this particular issue, and therefore that precious few seeming endorsements of one or the other of these possible retributivist desert objects reflects conscious choice. Nonetheless, once attention is drawn to the question a choice must be made—either between these two possible desert objects, or of some third.

If we just count noises, the dominant view, I think, is that they deserve to suffer. Moreover, there is reason to prefer this formulation, for if what they deserve is punishment, the argument for retributivism appears tautological or, at the least, uninformative: punishment of wrongdoers is justified by the fact that wrongdoers deserve to be punished. As Lawrence Davis argued some years ago in his

[10] K Greenawalt, 'Punishment' (1981) 74 Crim. L. & Criminology 343, 347. See also eg, RL Christopher, 'Deterring Retributivism: The Injustice of "Just Punishment"' (2002) 96 Northwestern U LR 843, 845 n.1 ("Though a precise definition of retributivism has proven elusive, stated most simply, the theory holds that punishment is justified solely because the person being punished deserves it."); M Moore, *Placing Blame* (1997) 91 ("Retributivism is a very straightforward theory of punishment: We are justified in punishing because and only because offenders deserve it."); H Bedau, 'Retribution and the Theory of Punishment' (1978) J Philosophy 601, 608 ("Retributivism without *desert*–the concept of punishment as something deserved by whoever is rightly made liable to it–is like *Hamlet* without the Prince of Denmark.").

[11] M Bagaric & K Amarasekara, 'The Errors of Retributivism' (2000) 24 Melbourne University LR 124, 127 (footnotes omitted); J Kleinig, *Punishment and Desert* (1973) 67 ("The principle that the wrongdoer deserves to suffer seems to accord with our deepest intuitions concerning justice."); RA Duff, 'Justice, Mercy, and Forgiveness' (1990) 9 Crim. Just. Ethics 51, 52 (noting "the central retributivist intuition that 'the guilty deserve to suffer"' and that punishment is "the infliction of suffering on the criminal").

aptly titled short essay "They Deserve to Suffer," retributivism avoids this particular objection if what I call its desert object is suffering, not punishment.[12] So let us start by construing the desert claim in this way: wrongdoers deserve to suffer. This is only a working assumption; we will return to consider other possibilities later. That is, we will consider different conceptions or specifications of the desert claim. For now, though we will work with this specification of the general retributivism desert claim: wrongdoers deserve to suffer (i.e., to endure some negative experiential state) on account of, and in proportion to, their blameworthy wrongdoing. To make clear that this is one possible *specification* of the desert claim, and not a purported *restatement*, let us call this the *desert-s claim*.

2.2. The Retributivist Intrinsic Good

Having thus refined the vague *desert claim* into the somewhat more precise *desert-s claim,* it remains to address the second question: what does the claim that wrongdoers deserve to suffer mean? Can the idea be equivalently expressed in other evaluative or deontic terms of which we think we have a surer grasp? 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 *retributivist 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-s claim,* theorists have supposed that the two are equivalent, but that the former is less obscure.[13] In 1993, Michael Moore, the foremost contemporary retributivist, himself endorsed the intrinsic good claim. Indeed, he allowed that the intrinsic good claim captures the core meaning of

[12] LH Davis, 'They Deserve to Suffer' (1972) 32 Analysis 136. Despite Davis's argument, I stand by the claim in text that few commentators, retributivist or otherwise, have consciously attended to possible variations on the desert claim. For a rare recent exception see DN Husak, 'Retribution in Criminal Theory' (2000) 37 San Diego L. Rev. 959, 972.

[13] *See eg,* Davis (n 12 above); T Hurka, 'The Common Structure of Virtue and Desert' (2001) 112 Ethics 6. The intrinsic good claim makes an early appearance in AC Ewing, *The Morality of Punishment* (1929) 14.

retributivism: "what is distinctively retributivist is the view that the guilty receiving their just deserts is an intrinsic good."[14]

As David Dolinko was among the first to recognize, Moore's acceptance of the intrinsic good claim poses a profound challenge to our dominant classificatory scheme.[15] 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![16]

That would be an absurd view, and Moore explicitly disavows it.[17] But then it might seem that retributivism is not interestingly different from other instrumentalist approaches to punishment, most of which also recognize that the consequences realization or pursuit of which justifies the imposition of criminal punishment are varied. After all, the principal reason to term the alternative to retributivism "consequentialism" rather than "utilitarianism" was precisely to deny that a consequentialist about punishment (what I am now calling an "instrumentalist" about punishment) must believe that utility constitutes the only metric of value. If instrumentalist justifications of punishment are, as a class, pluralist about value, then they comfortably encompass intrinsic-good retributivism. So it appears that retributivism just is that type of instrumentalism—call it "retributivist instrumentalism"—that recognizes intrinsic value in the suffering of wrongdoers.

You might think, however, that there is this difference between retributivist instrumentalism and all other instrumentalist justifications of punishment. For instrumentalists about punishment,

---

[14] MS Moore, 'Justifying Retributivism,' (1993) 27 Israel L. Rev. 15, 19 (emphasis omitted).
[15] D Dolinko, 'Retributivism, Consequentialism, and the Intrinsic Goodness of Punishment' (1997) 16 L. & Phil. 507.
[16] Ibid 513-14.
[17] *See eg,* Moore (n 14 above) 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 is justified by all the good consequences that punishment produces (or is reasonably expected to produce, or something of this sort). To be sure, punishment instrumentalists vary regarding what states of affairs *are* valuable. But it is a hallmark of instrumentalist accounts that all valuable states of affairs (whatever they might be) can help furnish justification for punishment. Yet retributivists often say that the offender's desert supplies both necessary and sufficient conditions for punishment.[18] We will focus on the supposed necessary condition shortly. Now just consider the sufficiency condition. If retributivism truly holds that the inflicting of deserved suffering is a sufficient justification for punishment, then it would seem to deny this tenet of instrumentalism.

In fact, the oft-stated retributivist contention that bringing about the good of deserved suffering is a sufficient condition for punishment cannot be taken at face value. If retributivists, like most folk, recognize a plurality of intrinsic goods, then they are apt to recognize a plurality of intrinsic bads as well. It would seem to follow 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 punishment. In fact, Moore has now granted 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.[19]

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, his pragmatic point being that the particular instrumentalist mechanisms frequently invoked and elaborated upon—

[18] See eg M Moore, *Placing Blame* (1997) 91 ("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.").
[19] Moore (n 14 above) 35.

deterrence, reform, moral education, and so on—are not themselves necessary conditions, alone or in combination, for punishment to be justified. If the sufficiency condition is understood in this way, it is entirely consistent with instrumentalist theories of punishment, in which event retributivism is simply a species of instrumentalist justifications for punishment. It is that type that recognizes the suffering of wrongdoers as intrinsically valuable—a valuable consequence that (for one reason or another) is particularly salient in justifying punishment.

Now, both Moore and Dolinko deny, albeit in different ways, that retributivism is best viewed simply as an instrumentalist theory that isolates deserved suffering as a particularly weighty value. Moore denies that retributivism *need* be assimilated to instrumentalism: in his view, there are instrumentalist and non-instrumentalist forms of retributivism, both of which are sound. Dolinko denies that retributivism *can* be assimilated to instrumentalism: in his view, retributivism is inescapably committed to claims that cannot be reconciled with instrumentalist modes of justification. Section 3 considers what routes are open to those, like Moore, who would favor a non-instrumentalist retributivism. Before considering whether a non-instrumentalist retributivism can be vindicated, however, we do well to first consider the plausibility of a genuine instrumentalist one. We do well, that is, to address Dolinko's challenge.

2.3. Is retributivist instrumentalism really a form of retributivism?

Recall Moore's assertion that the intrinsic good claim captures "what is distinctively retributivist."[20] "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."[21]

This is question begging, for a non-retributivist *cannot* endorse the intrinsic good claim if retributivism is *defined* by its endorsement of that claim. Then to endorse the intrinsic good claim would ipso facto to be a retributivist. So Dolinko must have in mind a different understanding of what

[20] See text accompanying n 14 above.
[21] Dolinko (n 15 above) 517.

retributivism necessarily is or maintains. Possibly, for example, he believes that retributivism is non-instrumentalist by definition. This is a not-uncommon view.[22] But it must be defended by argument.

Here is Dolinko's argument, or at least the assumption that undergirds his argument:

> [W] hat 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.[23]

This is a puzzling passage, for surely the intrinsic goodness of punishing the guilty (or of giving the guilty their deserved suffering) does play a role, for the retributivist, in helping to supply both rational justification and moral justification. Dolinko's use of the definite article seems to suggest, then, that a retributivist must believe that the intrinsic good that he emphasizes must be the *lone* rational or moral justification. Yet what grounds he has to saddle the retributivist with such a view is unstated. Dolinko provides no reason why a theorist who contends, say, that the good of deserved suffering is an always-available reason to justify punishment, and a particularly weighty reason at that, is barred from donning the retributivist mantle. That is, Dolinko provides no argument to preclude the possibility that the most plausible form of retributivism, and the version most widely embraced today, is a species of instrumentalism. It could be, in other words, that self-described contemporary retributivists *just are* those instrumentalists 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.[24]

---

[22] See eg RA Duff, *Punishment, Communication, and Community* (2001) 19; J Feinberg, 'What, If Anything, Justifies Legal Punishment? The Classic Debate' in J Feinberg & H Gross (eds) *Philosophy of Law* (5th ed 1995) 613, 613–17.
[23] Dolinko (n 15 above) 518.
[24] Possibly, however, retributivist instrumentalists are wrong about that. Instrumentalists who deny the intrinsic good claim (i.e., non-retributivist instrumentalists), 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

Even if Dolinko himself provides no persuasive explicit argument against retributivist instrumentalism, perhaps the germs of an argument can be found lurking in his disquiet. The worry, I think, is this: Supposing arguendo that retributivism is a form of instrumentalist justification, we should still want it to be the case that retributivist instrumentalism is not just any old instrumentalism. Instrumentalism about punishment is not committed to any particular view about the things of value that punishment should be designed to bring about; an instrumentalist's answers to that question are potentially as varied as are the possible views about value. Therefore, 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, it might seem that we would not be left with two species of instrumentalism—retributivist and non-retributivist—but dozens or scores rather than two, for every different axiological position would constitute its own punishment theory. For example, some instrumentalists about punishment who deny that suffering is ever intrinsically good might nonetheless be deontologists about morality who also 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. The worry, then, is not that the divide between "retributivist instrumentalism" and "non-retributivist instrumentalism" is false, but that it is arbitrary. If we were disposed to draw a binary classification within instrumentalist theories of punishment, we could just as well do it between welfarist and non-welfarist instrumentalism or deontological and non-deontological instrumentalism. But if all this is so, then perhaps we have reason to be skeptical of the move that got us here in the first place. It's one thing to swallow that retributivism could be a subset of instrumentalism. It's something else again to accept that it is no more salient or significant a subtype than many others. Of course, that could just be the way it is; I have thus far only tried to give voice to a

---

up backing doctrines and policies that closely approximate what a retributivist instrumentalist might advocate. See generally PH Robinson & JM Darley, 'The Utility of Desert' (1997) 91 Northwestern University LR 453.

worry, not to have burnished that worry into an argument. Still, reasoning of this sort can help explain the intuition that it is not enough that retributivism have a particular axiological commitment, but that it must display some sort of distinctive logical structure.

There is something to this concern. I will argue, though, that the adoption of the intrinsic good claim does in fact give retributivist instrumentalism a logical structure not shared by non-retributivist instrumentalism.

To understand the argument, we must first distinguish two different types of justification, what I will call "all-things-considered" and "tailored."[25] 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."[26]

As already discussed, the retributivist instrumentalist 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 Stanley Benn, it is "the deliberate imposition of suffering which is the feature needing justification."[27] 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

[25] The analysis to follow draws from my 'Punishment and Justification' (2008) 118 Ethics 258, 278-84.
[26] Cf n 9 above (invoking Feinberg's notion of a "desert basis").
[27] HLA Hart, *Punishment and Responsibility* (1968) 2 n.3.

of pain or other form of unpleasant treatment."[28] 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."[29] 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. In other words, they have 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.

Seeking an all-things-considered justification for punishment,[30] 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 instrumentalist 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

[28] C Finkelstein, 'Positivism and the Notion of an Offense,' (2000) 88 Cal. L. Rev. 335, 358. A tiny sampling of similar claims includes I Primoratz, *Justifying Legal Punishment* (1989) 7 ("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 . . . ."); R Wasserstrom, "Why Punish the Guilty?" in (1964) 20 Princeton University Magazine 14-19, reprinted in G Ezorsky (ed) *Philosophical Perspectives on Punishment* (1972) 328-41, 337 ("Punishment is an evil, an unpleasantness; it requires that someone suffer. Its infliction demands justification."); N Lacey, *State Punishment: Political Principles and Community Values* (1988) 13 ("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.").
[29] RW Burgh, 'Do the Guilty Deserve Punishment?' (1982) 79 J. Phil. 193, 193.
[30] *See eg* Dolinko (n 15 above) 521 (inquiring whether retributivism can "rationally justify the practice of punishment in the face of its staggering costs").

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 instrumentalist 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.

Put another way, once we recognize the centrality, in the philosophical literature, of the search for tailored justifications for punishment, then we can agree with Moore that "what is distinctively retributivist is the view that the guilty receiving their just deserts is an intrinsic good" without abandoning the intuition that there should remain something importantly and revealingly distinct about retributivist and instrumentalist accounts of the justifiability of punishment. Even insofar as retributivism cashes out as retributivist instrumentalism, it would remain in some respect structurally distinct from non-retributivist instrumentalism.

3. Non-instrumentalist Retributivism

Let us take stock. Retributivism is constituted by its acceptance of *the desert claim*: punishment is justified in terms of the wrongdoer's ill-desert. The desert claim is consistent with potentially many specifications of the desert object. The dominant specification holds (to a first approximation) that what wrongdoers deserve is to suffer on account of their blameworthy wrongdoing. This is *the desert-s claim*. The desert-s claim is equivalent to *the retributive intrinsic good claim*: it is intrinsically valuable that wrongdoers suffer on account of their blameworthy wrongdoing. Embrace of the intrinsic good claim, by Moore and other retributivists, converts retributivism into an instrumentalist justification for punishment. That is, to justify punishment on the strength of the retributive intrinsic good claim is to advance an instrumentalist justification for punishment, and not to advance an alternative or

competitor to the instrumentalist justifications. However, given the importance to moral reasoning of the distinction between tailored and all-things-considered justifications, there remains good reason to single out retributivist instrumentalism as an especially meaningful subtype.

What we have thus far been calling retributivist instrumentalism could also be termed instrumentalist retributivism; these are alternative descriptions of the same idea.[31] The final question is whether instrumentalist retributivism just is retributivism today or whether there are, in addition, one or more tenable forms of non-instrumentalist retributivism. If there are, a proponent of such forms of retributivism might maintain, modestly, that both are forms of retributivism or, more aggressively, that so-called instrumentalist retributivism isn't a genuine form of retributivism at all. This latter position would be, in essence, to agree with Dolinko, but from the other side of the aisle. I have already provided reason to deny that, if there is a viable or vibrant form of non-instrumentalist retributivism, it alone is entitled to the label retributivism: many persons who would self-identify either as retributivist or non-retributivist seem to believe that acceptance or denial of the retributivist intrinsic good claim is of central importance to proper classification of their views, and retributivist instrumentalism supplies tailored justification in a fashion that other forms of instrumentalism do not. Therefore, I will suppose that non-instrumentalist retributivists need not deny either (1) that instrumentalists about punishment differ on the question of whether the suffering of a wrongdoer is an intrinsic good or (2) that they share something important in common with those who answer that question in the affirmative, but rather that they need insist only (3) 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 instrumentalism.

[31] Which is the modificand, and which the modifier, depends upon what is being taxonomized. When we speak of justifications for punishment, I think it most perspicuous to distinguish, at the top level of the taxonomic hierarchy, instrumentalist from non-instrumentalist theories. The retributivist subclass of instrumentalism is naturally denominated retributivist instrumentalism. But if we are inquiring into the possible or actual subtypes of retributivism, then what we call retributivist instrumentalism in the first discursive context becomes instrumentalist retributivism in this second. (Compare: as an American, Malcolm X was notable for being a Muslim American; on the hajj, he was more likely seen as an American Muslim.)

The claim that all contemporary retributivists and instrumentalists 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. 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."[32] So we ought to expect that a genuine and plausible non-instrumentalist retributivism can be articulated (whether or not it can ultimately withstand critical scrutiny). But the nature of this non-instrumentalist retributivism is more elusive than is its instrumentalist sibling. This final section, accordingly, canvasses a variety of ways that a retributivist might possibly escape the reduction of his or her view to instrumentalist retributivism. My goal will not be to reach a verdict on non-instrumentalist retributivism but to identify the range of avenues open to approaches of this sort and to offer at least a modicum of critical analysis—trying to show which avenues are more or less promising and why. Very simply, it is clear that at least some theorists of a retributivist bent will resist equating the desert claim with the intrinsic good claim, but it is not yet entirely clear what are the best grounds for blocking that translation. The remainder of this paper aims to make a start—but only a start—at examining that question. To emphasize: we are not looking now for just any alternative to instrumentalist justifications for punishment, but for an alternative that is recognizably retributivist in bearing a sufficiently intimate relationship to the notion of an offender's ill-desert.[33] Non-instrumentalist justificatory accounts more properly described, for example, as expressive or aretaic would not fit the bill.

To preview: this section identifies four reasonably familiar accounts of the features of a putatively retributivist justification of punishment that would serve to distinguish it from a retributivist

[32] G Bassham, 'Justice Scalia's Equitable Constitution' (2006) 33 J.C & U.L. 143, 154.
[33] To be sure, this is to accept that ill-desert is somehow constitutive of retributivism. This is a contestable proposition, but not widely contested in fact and therefore something we should not abandon without good reason.

instrumentalism (i.e., from an instrumentalist retributivism). To a first approximation, these are the claims that non-instrumentalist retributivism is distinctive in: (1) precluding punishment of the innocent; (2) conceiving of punishment as a dictate of justice; (3) providing that we have a duty or obligation to punish wrongdoers, and not merely that such punishment is permissible or justified; and (4) denying that the rightness of punishment is reducible to any claims about the good or the valuable. I will argue (admittedly, in a fairly conclusory way) that the first of these three candidate grounds for marking a genuinely non-instrumentalist retributivism is not tenable, and that the second and third are unpromising. I will then conclude that the fourth is the most promising of these proposed routes. Because it is beyond the scope of this chapter to reach a verdict on just how plausible that route is, let alone whether it is correct, this section rests with raising some difficulties for this approach and offers some thoughts about its prospects. The most original and important contribution of this section will be the claim that the difference between an instrumentalist and a non-instrumentalist retributivism is most perspicuously reduced simply to a matter that we have already touched briefly upon: how best to specify the retributivist desert claim.

3.1. Negative retributivism

Recall the frequent retributivist claim that an offender's desert furnishes both a necessary and sufficient justification for punishment. We have seen that if the desert claim is translated, by steps, into the intrinsic good claim, then the sufficiency condition does not seem maintainable. Let us now consider the necessity condition. One proposal is that non-instrumentalist retributivism is distinct from any form of instrumentalism (including retributivist instrumentalism) not because of what justifies punishment, but by what limits it.[34] Instrumentalist theories license punishment in the absence, or in excess, of an offender's desert; non-instrumentalist retributivism would bar this.

[34] This understanding of retributivism was identified and rightly criticized in J Cottingham, 'Varieties of Retribution' (1979) 29 Phil. Q. 238, 240-41, but it still pops up from time to time.

Before assessing this "necessary condition" claim, we should note that it is more modest than 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 instrumentalist justifications for punishment as the retributivist sufficiency condition did not, for full-blooded consequentialists (about morality) might be thought compelled to reject the necessity claim even in this qualified form.

This is demonstrably not a plausible basis for marking out non-instrumentalist retributivism. Not only *can* instrumentalists accept that the ill-desert of an individual supplies a necessary condition on the morally permissible imposition of punishment, but many do.

Instrumentalist 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 offenders on the non-consequential ground that he or she violated a legal command.[35] 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."[36]

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 clearly

[35] J Rawls, 'Two Concepts of Rules' (1955) 64 Phil. Rev. 3.
[36] HLA Hart, *Punishment and Responsibility* (1968) 8-12.

insisted that justice demands that punishment not be imposed on someone known not to deserve the suffering imposed[37]—a moral view that John Mackie would soon dub "negative" retributivism,[38] and that Antony Duff would call, perhaps even more aptly, "side-constrained consequentialism."[39] Today, instrumentalists about punishment 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. (I do not claim that such acceptance is somehow compelled, but only that it is widespread.) I will call an instrumentalist theory that is limited in this way "desert-constrained instrumentalism" to foreground clearly the nature of the side-constraint—namely, knowingly punishing in the absence, or in excess, of ill-desert.

How could an instrumentalist adopt such a view? In at least two ways. First, as several moral philosophers have recently argued, putatively deontic side constraints might be accommodated within a full-blooded consequentialist ethics by the recognition of agent-relative value.[40] I am myself somewhat skeptical of this line of argument, but this cannot be the place for a critical evaluation. Suffice to say that the possibility cannot be ruled out.

Much more significantly, instrumentalists about punishment need not be full-blooded consequentialists. That, of course, is the very point of renaming as "instrumentalist" what had been deceptively captioned "consequentialist" justifications for punishment.[41] Because instrumentalists about punishments need not be consequentialists about ethics, they confront no contradiction in

---

[37] I say that neither "clearly" insisted upon this because Hart's signals on this particular score are famously ambiguous. Although he often defended the principle of "retribution in distribution" on the grounds that it supplied people with the seemingly welfarist benefits of living under a "choosing system," he also described it in terms of "justice" and "fairness" that seem, in context, non-utilitarian. Particularly compare, in ibid, ch 2 'Legal Responsibility and Excuses' with ch 3 'Murder and the Principles of Punishment: England and the United States.'

[38] JL Mackie, 'Morality and the Retributive Emotions' (1982) 1 Criminal Justice Ethics 3, 3.

[39] RA Duff, *Punishment, Communication, and Community* (2001) 11. The locus classicus of the notion of side-constraints on consequentialism is, of course, R Nozick, *Anarchy, State and Utopia* (1974) 28-35.

[40] Recent defenses of this view include M Smith, 'Two Kinds of Consequentialism' (2009) 19 Philosophical Issues 257; DW Portmore, 'Consequentializing' (2009) 4 Philosophy Compass 329; M Peterson, 'A Royal Road to Consequentialism' [2009] Ethical Theory and Moral Practice. For a criticism, see Mark Schroeder, 'Teleology, Agent-Relative Value, and "Good"' (2007) 117 Ethics 265.

[41] This is a central theme of V Tadros, *The Moral Foundations of Criminal Law* (forthcoming) ch __.

contending that punishment is justified by the goods that the practice produces (including, for retributivist instrumentalists, the good of meting out deserved suffering) but only so long as we do not punish persons believed to lack personal ill-desert.

3.2. Retributive justice

A second possible route to non-instrumentalist retributivism starts with the observation that retributivism is routinely described as a theory of justice: It maintains that justice demands that offenders be given their ill-deserts. As Michael Moore put it in a much-quoted passage, the desert of the offender "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.[42] Call this *the retributive justice claim.* It is not equivalent, so the argument must go, to the retributivist instrumentalist'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.

Exactly how characterizing retributivism as a theory of justice paves the way toward non-instrumentalist retributivism is not at all clear. The proponent of such a contention must first explain what it means for a moral norm to be a norm *of justice.* Plainly, norms of justice are not categorically more stringent than other moral norms. The moral prohibition on killing innocent non-aggressors is stringent indeed, yet it is rarely if ever characterized as an obligation of justice.

On one common view, norms of justice are those moral norms that possess a certain sort of public or political character, justice being "the first virtue of social institutions," in Rawls's language.[43]

[42] MS Moore, 'The Moral Worth of Retribution' in FD Schoeman (ed), *Responsibility, Character, and the Emotions* (1987) 179, 182.
[43] J Rawls, *A Theory of Justice* (rev. ed. 1999) 3.

This is to put things very vaguely, but whatever its details, this understanding of justice is unlikely to help distinguish non-instrumentalist retributivism from competing accounts, for criminal punishment is public or political no matter what its justifications. Another widespread conception distinguishes norms of justice from other moral norms, not in virtue of their strength or their political character, but only in terms of their subject matter. As John Gardner and Francois Tanguay-Renaud put it in this volume, justice is the name for moral norms concerned with the allocation of benefits and burdens.[44] I do not know if this is so. But if something like this is true, then invocation of retributive justice would again not seem, by itself, to provide grounds for a distinct non-instrumentalist justification for punishment. If there *is* a non-instrumentalist retributivism then it might very well be a justification for punishment properly classed as a dictate of justice—as, possibly, instrumentalist justifications of punishment (including retributivist instrumentalism) are not. But the justice label is not itself doing the work. What will do the work is whatever it is that makes the non-instrumentalist retributivism true.

3.3. The duty to punish

Consider, again, the passage just quoted from Moore. We took the idea that non-instrumentalist retributivism might be distinctive in virtue of its being a dictate of *justice* from the third sentence. But the first two suggest a different idea, and one with somewhat greater promise: non-instrumentalist retributivism might be distinguishable from its instrumentalist sibling on the grounds that the former, but not the latter, maintains that satisfaction of the offender's ill-desert grounds a duty to punish, and not merely that it renders punishment morally justifiable.[45]

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 disputed—I am aware of no contemporary

[44] [EDITORS: please cite to Gardner & Tanguay-Renaud]
[45] For discussion of retributivist views that emphasize the duty to punish see MT Cahill, 'Retributive Justice in the Real World' (2007) 85 Wash U LR 815, 825, 826-29.

philosopher or criminal law theorist who advances such an extreme position. Accordingly, a far more plausible construal of the deontic duty claim holds that we have a pro tanto obligation to realize the wrongdoer's ill-desert. The questions, then, are whether many contemporary retributivists endorse such a claim, and in a manner that instrumentalists cannot endorse as well. Let's consider two possible ways to cash out the retributivist duty claim: in terms of protected reasons and in terms of agent-relativity.

As a pluralist, the retributivist instrumentalist 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. Retributivists who believe we have a duty to punish might believe that 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, our supposed duty to punish might reflect 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.[46] I imagine this is possible, but it does not appear a particularly promising route to non-instrumentalist retributivism, for it is not at all apparent why instrumentalists must find recourse to protected reasons uncongenial.

Consider, then, a second possible way to make sense of the retributivist duty claim—in terms of agent-relativity. Imagine this proposal from Cruel Dictator to Just Republic: "iff you impose punishment of type and magnitude n (call such punishment "Pn") on Non-Wrongdoer1 (call him "NW1"), whom you and I know to be innocent, then I will refrain from imposing Pn on NW2 and on NW3, both of whom we

[46] On protected reasons, see J Raz, *Practical Reasons and Norms* (1999) 191.

know to be innocent." May Just Republic accept this proposal, thereby inflicting Pn on NW1? Presumably no. That's why "negative retributivism" seems to capture a genuine deontic duty (even if defeasible at a threshold, and subject to the possibility, already mentioned, that such duties can be consequentialized).

Is there a similar case to be made for an agent-relative duty to punish? Consider now this proposal from Benevolent Monarch to Just Republic: "iff you refrain from imposing Pn on Wrongdoer1 (W1), whom you and I know to be guilty, then I will impose Pn on W2 and W3, both of whom we know to be guilty." (Testimony or evidence from W1 is needed to secure a conviction of W2 and W3.) May Just Republic accept this proposal, thereby not inflicting Pn on W1? If the state were under an agent-relative duty *to* punish that roughly mirrored its agent-relative duty *not* to punish (i.e., if there were a side-constraint on *not* punishing comparable to the side-constraint many people believe applies to punishing), then the answer must be no. My guess, though, is that a great many theorists who self-identify as retributivists—even those would resist being characterized as retributivist instrumentalists—would say yes. If so, that shows that they do not recognize an agent-relative duty to punish wrongdoers. To be clear, this is not an argument that the contention that we have a duty to punish wrongdoers, if sound, would not constitute a non-instrumentalist retributivism. Nor is it an argument that such a contention cannot in fact be maintained: readers who respond to this hypothetical differently than I anticipate should indeed endeavor to develop the supporting arguments. I mean only to cast doubt that this is the most promising route for would-be non-instrumentalist retributivists to pursue.[47]

3.4. The rightness of punishing

If that is so, we come to this alternative: non-instrumentalist retributivism might be best constituted not in terms of the state's duty to punish but rather in terms of the rightness of its doing so,

[47] Moreover, and again, the agent-relative construal of the retributive duty claim is arguably consequentializable, thus amenable to an instrumentalist justification for punishment, if there exists agent-relative value.

where "rightness" can be understood in terms of kindred notions like "oughtness" and, yet more commonly these days, reasons. More particularly, non-instrumentalist retributivists might wish to argue that punishment is justified because doing so is right—something we have reason, or ought, to do—and where its rightness is not derivative of its being valuable. Of course, instrumentalist retributivists may also justify punishment as "right." They would say that it's right because (and insofar as) it promotes the good. So this path toward a truly non-instrumentalist retributivism depends not on the supposed rightness of punishment but on the claim that its rightness does not derive from its being valuable.

How could punishment be right but not because it is valuable? One possibility is that punishment is right and not valuable. This seems unpromising, and not only because, in the estimation of one commentator, "[t]o be a retributivist is to recognize that deserved punishment is an intrinsic good."[48] Although there may be some actions that are right yet wholly without value,[49] surely those are by far the exception. More promising is the claim that, although punishment is valuable in addition to being right, the value of punishment is not what makes it right. It could be, for example, that the good is defined in terms of the right—that that which is valuable has value in virtue of our having reasons of certain sorts with respect to it. Or, perhaps reasons and values can have the same ground without either being derived from the other. There are other possibilities too.[50]

Perhaps this is the route for non-instrumentalist retributivists to take. But they confront this challenge: to explain the ground of the right in a way that is properly linked to the notion of an offender's ill-desert. How can it be, if what a wrongdoer deserves is to suffer, that the rightness of

[48] L Zaibert, *Punishment and Retribution* (2006) 214.
[49] As Jonathan Dancy explains, "An action can be one's duty even though doing it has no value and its being done generates nothing of value. Standard examples here are of trivial duties." J Dancy, 'Should We Pass the Buck?' in A O'Hear (ed) *Philosophy, The Good, The True and The Beautiful* (2000) 159, 168. Presumably if this is true of duty, the same would hold true, a fortiori, of what is right.
[50] These and other possibilities are trenchantly explored in Dancy, which challenges the first possibility on offer (the "buck-passing" view). The buck-passing view is also criticized in W Rabinowicz & T Rønnow-Rasmussen, 'The Strike of the Demon: On Fitting Pro-attitudes and Value' (2004) 114 Ethics 391.

engaging in actions designed to bring such suffering about does not derive from the goodness of his suffering?

The answer to this challenge, I think, is this: non-instrumentalist retributivists who would justify punishment as right but not because of its instrumental value must deny that what a wrongdoer deserves is to suffer. Recall that we accepted as constitutive of retributivism what we called *the desert claim:* punishment is justified by the offender's ill-desert. What led us to retributivist instrumentalism was our (provisional) endorsement of a particular specification of this desert claim that we termed *the desert-s claim:* wrongdoers deserve to suffer (i.e., to endure some negative experiential state) on account of, and in proportion to, their blameworthy wrongdoing. Non-instrumentalist retributivists must reject that specification of the desert claim. In particular, they must articulate and defend a competing specification of the wrongdoers' desert object that is not expressible without reference to actions by a responsive agent. The most obvious possibility is for them to return, despite the worry that provoked Lawrence Davis to endorse the *desert-s claim* nearly forty years ago, to the claim (*the desert-p claim*) that what wrongdoers deserve is to be punished. But this is not their only option. Equally non-instrumentalist is Duff's claim (*the desert-c claim*) that what they deserve is to be censured.[51] In principle, there exist any number of other possibilities[52]—just as *the desert-s claim* may not be the only specification available to instrumentalist retributivists. The critical point is that a choice needs to be made between the possibilities that what wrongdoers deserve is a response of some kind from the group (be it the community or the state), or that what they deserve is a state of affairs that is itself describable without reference to any human agency that may or may not be charged with causing that state of affairs to obtain. The question that divides non-instrumentalist from instrumentalist

[51] Eg RA Duff, *Punishment, Communication, and Community* (2001) 27-30.
[52] For example, David Enoch and Larry Sager have combined to suggest to me a possibility that approximates, but may improve upon, *desert-p*: wrongdoers deserve to be made to suffer by, or under the aegis of, the sovereign.

retributivism, on this view, is the answer each gives to the question of whether the retributivist desert object is response-dependent or response-independent.[53]

Put another way, Husak might have been too quick in asserting that "retributive beliefs only require that culpable wrongdoers be given their just deserts by being made to suffer (or to receive a hardship or deprivation). These beliefs do not require that culpable wrongdoers be given their just deserts by being made to suffer by the state through the imposition of punishment."[54] It might be that some retributive beliefs *do* require not merely that wrongdoers experience some state of affairs, but that they be met with some specific sort of *response*. In short, the battle between the two kinds of retributivism—instrumentalist and non-instrumentalist—is best fought out in the space of specifying the wrongdoers' supposed desert object. Specification of the desert claim is not a sideshow. It is the whole game, for the best way to establish that the rightness of punishing wrongdoers is not derivative of the goods that punishment produces might be just to articulate and defend a conception of an offender's desert object in relational or "responsive" terms. (This is reflected in the following table.)

Specification of the desert object	Equivalent formulations (to a first approximation)	Instrumentalist or non-instrumentalist
Wrongdoers deserve to suffer	It is intrinsically good that wrongdoers suffer	instrumentalist
Wrongdoers deserve to be punished	It is right that wrongdoers are punished; we ought to punish wrongdoers	Non-instrumentalist
Wrongdoers deserve to be censured	It is right that wrongdoers are censured; we ought to censure wrongdoers	Non-instrumentalist

3.5. Contingent and conceptual instrumentalism

[53] Although Feinberg did not offer a pithy label for what I am calling the "desert object," he did consistently refer to the "modes [or kinds] of treatment" that people could deserve. Were it constitutive of desert objects that they must be modes of *treatment,* then instrumentalist retributivism would be doomed from the outset. Although Feinberg's own language seems to assume that is so, he provides no argument for it and I am aware of none.
[54] DN Husak, 'Retribution in Criminal Theory' (2000) 37 San Diego L. Rev. 959, 972.

I would like now to pursue this idea from one other direction. Several theorists with retributive sympathies who have acknowledged the plausibility of recasting retributivism in instrumentalist terms have nonetheless sought to resist what might appear to be the marginalization of retributivism as a distinct punishment theory by proposing that retributivist instrumentalism invokes a very different type of consequence than does non-retributivist instrumentalism. George Fletcher pursued this line of argument when distinguishing between "factually" and "conceptually" consequentialist theories. [55] Others may prefer to speak of "contingent" and "intrinsic" consequentialism. 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.[56]

Theorists who pursue this line for distinguishing retributivism from instrumentalism appear to concede (at least arguendo) that all theorists who would justify punishment do so by reference to its consequences. They contend that the important question is whether the consequences that do the justificatory work are caused by punishment contingently or as a matter of the logic of the concept. This is not the distinction I am proposing, but close enough to invite comment on why it isn't.

Take a retributivist who embraces the *desert-s claim.* By choosing this specification of the retributivist desert object, she is, in my terms, an instrumentalist retributivist. If what I call non-instrumentalist retributivism were equivalent to what others might be disposed to characterize as the

[55] GP Fletcher, 'Punishment and Responsibility' in D Patterson (ed) *The Blackwell Companion to The Philosophy of Law and Legal Theory* (1996) 514, 516.
[56] Zaibert (n 48 above) 133. See also RL Christopher, 'Deterring Retributivism: The Injustice of "Just Punishment" (2002) 96 Nw. U. L. Rev. 843.

retributivist subtype of "intrinsic" (or "conceptual") consequentialism,[57] then insofar as a justification of punishment relies upon desert-s, it would be, to that extent, contingently consequentialist, not intrinsically so. But whether acceptance of the *desert-s claim* underwrites a contingently consequentialist justification for punishment is not clear: it depends on what we consider "the very act of punishment."

If punishment is *defined* (to a first approximation) as *the infliction* of suffering on (actual or supposed) wrongdoers in response to their (actual or supposed) offense, then the wrongdoer's suffering is an intrinsic consequence of each act of punishment. In contrast, if punishment is defined, say, as a variety of forms of conduct *designed* or *intended* to be burdensome, then whether such acts do in fact cause suffering is a contingent question. Of course, these are far from the only definitional possibilities. The important point, though, is that whether a retributivist account that would justify punishment by (in significant part) its tendency to bring about the state of deserved suffering belongs to contingent or intrinsic consequentialism all depends on how we define punishment. But that is just to lead us back into the definitional debates that once occupied the field and that most philosophers of the criminal law, I submit, are very pleased we have escaped.

We do not, I think, want to have our best understanding of different justifications of punishment—here, whether a particular justificatory account belongs to contingent or intrinsic consequentialism—depend upon more precise and contestable specifications of what punishment is. What we may want, though, is for them to depend on competing specifications of what wrongdoers deserve. If the desert object is response-dependent, then punishment (of some form) is itself what wrongdoers deserve. Retributivists of a non-instrumentalist bent who were attracted to the distinction between intrinsic and contingent consequences might, then, have things close to backwards. We should

[57] I am assuming that intrinsic consequentialism encompasses theories that are not retributivist.

not ask whether the desert object (that which, for retributivists, punishment is designed to achieve or bring about) is an "intrinsic consequence" of punishment, but whether a punitive response is an intrinsic feature of the desert object.

CONCLUSION

For several generations now, philosophers of the criminal law have routinely divided the potentially messy universe of justificatory accounts of criminal punishment into just two broad classes: retributivist and consequentialist—classes that are I have chosen to render as retributivist and *instrumentalist* to foreground that adherents of this latter position need not endorse consequentialism as a comprehensive moral theory. Broadly speaking, retributivists have contended that punishment is justified by the fact that an offender deserves it. Instrumentalists argue that punishment is justified by the net good consequences that the practice produces, paradigmatically but not necessarily exclusively, in terms of deterring anti-social aggression.

But what is the "it" that, on the retributivist account, wrongdoers deserve? One answer—possibly the most common—is that they deserve to suffer. If this is so, then it seems natural to refashion the retributive desert claim into the claim that that it is intrinsically valuable that wrongdoers suffer on account of their blameworthy wrongdoing. This latter claim is said by some retributivists to be equivalent to the desert claim but somewhat less mysterious and therefore more plausible to the uncommitted. But the retributive embrace of the intrinsic good claim seems to make retributivism amenable to an instrumentalist structure of justification. Suddenly, retributivism looks like a type of instrumentalism, not a competitor.

Those who would resist this conclusion have two moves open to them. First, they could deny that retributivist instrumentalism deserves the retributivist label at all. I have argued that this is too

dogmatic. This claim on which this account rests—namely, that it is intrinsically valuable for a wrongdoer to suffer—is recognizably within a family of classically retributivist ideas. The second avenue, then, is to explicate a distinct alternative retributivism not similarly amenable to instrumentalist justification. Success in this pursuit would leave us with two kinds of retributivism, not one.

How best can a non-instrumentalist retributivism be vindicated? Retributivists who would resist instrumentalism have been tempted down many paths, including those that emphasize: their commitment to the impermissibility of punishing those believed not to be wrongdoers; retributivism's supposed status as a form of justice; and the inability of instrumentalist justifications to make sense of our supposed duty to punish. I have offered reasons to think that none of these paths holds great promise, and that the most promising route is what many readers are likely to have thought most obvious in the first place: to make use of the proposition that it is right for us to punish wrongdoers—in the sense that we ought to do so, or have moral reason to do so—on grounds not derived from the value of punishing them or the value of states of affairs caused by our punishing them. But just how to defend a view of this sort that makes it appear as more than ipse dixit remains the great retributivist challenge. I have argued that the debate between what are now two kinds of retributivism might be most perspicuously reformulated as a debate over just what it is that wrongdoers deserve. If this is correct, then efforts to articulate and defend different specifications of wrongdoers' desert object—precisely what it is that they deserve in virtue of their wrongdoing—will repay careful philosophical attention.