

University of Pennsylvania Carey Law School

Penn Law: Legal Scholarship Repository

Faculty Scholarship at Penn Law

2008

Punishment and Justification

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., "Punishment and Justification" (2008). *Faculty Scholarship at Penn Law*. 2348.
https://scholarship.law.upenn.edu/faculty_scholarship/2348

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.

Punishment and Justification

Mitchell N. Berman*

INTRODUCTION

It is sometimes suggested that philosophical discussions of punishment are plagued by radical dissensus.¹ But this is an overstatement. On at least one matter, there is widespread agreement, indeed a consensus approaching unanimity: Punishment stands in need of justification. Very often, this principle constitutes the theorist's first premise. It is, in any event, a proposition upon which possibly all philosophers of the criminal law are agreed.²

* Bernard J. Ward Centennial Professor in Law, the University of Texas at Austin. mberman@law.utexas.edu. For helpful comments on previous drafts, and for profitable conversations, I am grateful to Larry Alexander, John Deigh, David Dolinko, Strefan Fauble, Les Green, Tom Green, Mark Greenberg, Don Herzog, Douglas Husak, Larry Sager, Scott Shapiro, Peter Westen, and to two anonymous reviewers. I am also indebted to workshop participants at the law schools of Columbia University and the University of Michigan, and at the Second Annual International Congress on Law and Philosophy at the National Autonomous University of Mexico.

¹ E.g., Nicola Lacey, *State Punishment: Political Principles and Community Values* (1988), p.1.

² See, e.g., Claire Finkelstein, "Positivism and the Notion of an Offense," *California Law Review* 88 (2000): 335-394, p.358 ("Philosophical discussions of punishment often begin with the observation that because it involves the infliction of pain or other form of unpleasant treatment, punishment stands in need of justification.") (citing works by Feinberg, Benn, Flew, and Hart); R.A. Duff, *Trials and Punishments* (1986), p.4 ("It is agreed that a system of criminal punishment stands in need of some strenuous and persuasive justification"); Lawrence Crocker, "The Upper Limit of Just Punishment," *Emory Law Journal* 41 (1992): 1059-1110, p.1063 ("Essays on punishment are often promoted as providing a justification for punishment, no doubt because punishment is obviously very much in need of a justification."); John Kleinig, *Punishment and Desert* (1973), p.1 ("Punishment is said to stand in need of justification."); Ted Honderich, *Punishment: The Supposed Justifications* (1969), p.8 ("one cannot but regard

Of course, proposed responses to the need for justification vary more widely than the standard distinction between consequentialist and retributivist justifications might imply. Consequentialist accounts differ in emphasizing deterrence or incapacitation or norm reinforcement. Retributivist theories are even more diverse. Indeed, there is so little consensus on what retributivism claims that some would define it merely as non-consequentialism.³

Notwithstanding the variation within each broad camp, consequentialist and retributivist approaches face well-known objections. Briefly, consequentialists are thought misguided for denying the intuitive appeal of the idea that wrongdoers deserve to suffer, while their opposition to a practice of scapegoating is condemned as weak and contingent. Retributivists are said to be unable to make clear either why wrongdoers deserve to suffer or why it is permissible for a state institution to inflict suffering even if deserved. They are also deplored as savage or barbaric.

In light of these and other critiques, many scholars have expressed a preference for “mixed” or “hybrid” theories of the justifiability of criminal punishment.⁴ Admittedly, to attribute widespread attraction to hybrid accounts runs against prevailing fashion given the frequently voiced contention that retributivism, albeit given up for dead a generation or more

punishment as in need of justification”); C.L. Ten, *Crime, Guilt, and Punishment* (1987), p.3.

³ See, e.g., Joel Feinberg, “What, If Anything, Justifies Legal Punishment? The Classic Debate,” in *Philosophy of Law* eds. Joel Feinberg & Hyman Gross (5th ed. 1995): 613-617, p. 612. This position is not, however, generally well received. See, e.g., John Cottingham, “Varieties of Retribution,” *Philosophical Quarterly* 29 (1979): 238-46; R.A. Duff & D. Garland, “Introduction: Thinking About Punishment,” in *A Reader on Punishment* eds. Duff & Garland (Oxford: Oxford University Press, 1994): 1-34, p. 8.

⁴ One extensive yet partial list of theories plausibly characterized as hybrid appears in Crocker, p. 1062 n.8.

ago, is the dominant theory of punishment today.⁵ But that is hard to credit.⁶ It is not incidental that the lion's share of declarations announcing retributivism's ascendance appear as prologues to scathing critique. With others, then, I would suppose that at least a plurality of contemporary Anglo-American criminal law theorists accept some form of hybrid account.⁷

The most famous accounts customarily classified as hybrid belonged to John Rawls and H.L.A. Hart. 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.⁸ 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

⁵ See, e.g., David Dolinko, "Three Mistakes of Retributivism," *UCLA Law Review* 39 (1992): 1623-1657, p. 1623 ("It is widely acknowledged that retributivism . . . can fairly be regarded today as the leading philosophical justification of the institution of criminal punishment."); Mirko Bagaric & Kumar Amaraskara, "The Errors of Retributivism," *Melbourne University Law Review* 24 (2000): 124-189, p.126. Citations to works proclaiming a retributivist renaissance are collected in Russell L. Christopher, "Deterring Retributivism: The Injustice of "Just" Punishment," *Northwestern University Law Review* 96 (2002): 843-976, pp. 845-847 & nn.2-12.

⁶ I am not the first to express doubts that retributivism is the dominant theory of punishment among contemporary criminal law theorists. See, e.g., Kyron Huigens, "The Dead End of Deterrence, and Beyond," *William & Mary Law Review* 41 (2000): 943-1036, p. 955.

⁷ A similar, but stronger, judgment is expressed in Douglas N. Husak, "Why Punish the Deserving?," *Nous* 26 (1992): 447-464, p. 452 (asserting that "almost all philosophers appreciate the need to construct hybrid syntheses of retributive and consequentialist elements"). See also, e.g., A. John Simmons et al., "Introduction," in *Punishment: A Philosophy & Public Affairs Reader* eds. A. John Simmons et al. (1994): vii-xii, pp. viii-ix.

⁸ John Rawls, "Two Concepts of Rules," *Philosophical Review* 64 (1955): 3-32.

distribution” that permits imposition of punishment only on “an offender for an offense.”⁹ A popular cousin view to Hart’s, variously dubbed “negative”¹⁰ or “weak”¹¹ retributivism, holds that punishment is justified by its good consequences – consequences realized most especially through the mechanisms of general and special deterrence, incapacitation, and a variety of expressive or communicative functions – but ought to be limited by the desert of the offender.

But all of these views have been criticized for marginalizing retributivism relative to consequentialism.¹² This is plainly true of retributivism’s “negative” variant, which treats a wrongdoer’s desert as a side-constraint on pursuit of consequentialist aims and, therefore, is more aptly described as “side-constrained consequentialism”¹³ than as any form of retributivism. Yet the marginalization of retributivism is even more extreme in the Hartian and Rawlsian accounts, for which it can reasonably be doubted whether there exists any component that genuinely warrants the retributivist label.¹⁴ Indeed, the ballyhooed retributivist revival is better viewed, I

⁹ H.L.A. Hart, *Punishment and Responsibility* (1968), pp. 8-12.

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

¹¹ Hart, *Punishment and Responsibility*, pp. 230-37; Larry Alexander, “Retributivism and the Inadvertent Punishment of the Innocent,” *Law and Philosophy* 2 (1983): 232-246, pp. 238-39; George Schedler, “Can Retributivists Support Legal Punishment,” *Monist* 63 (1980): 185-198, p. 186.

¹² See David Wood, “Retribution, Crime Reduction and the Justification of Punishment,” *Oxford Journal of Legal Studies* 22 (2002): 301-21.

¹³ R.A. Duff, *Punishment, Communication, and Community* (Oxford: Oxford University Press 2001), p. 11.

¹⁴ For doubts that Hart’s account is fairly described as any sort of compromise with retributivism, see Cottingham, p. 241. It seems to me that, insofar as the basically

think, as a reaction against the marginalization of retributivism in even purportedly hybrid accounts than as an effort to turn the tables by marginalizing, let alone jettisoning, consequentialism in return. In other words, theorists have sought to replace hierarchical or “split-level”¹⁵ accounts with accounts that give consequentialism and retributivism something closer to co-top billing. Yet the straightforward contention that desert and good consequences are jointly necessary to the justification of punishment – what one scholar has recently dismissed as a “mere conjunction” approach – seems ad hoc or poorly supported.¹⁶ More satisfying might be an “integrated dualist” account – a theory that assigns distinct and co-equal roles to retributivist and consequentialist reasoning within the theory’s structural logic. However, no such account has yet appeared, and the possibility of such an appearance has been doubted.¹⁷

consequentialist accounts of Hart and Rawls are properly deemed “hybrids,” they are hybrids only with the legalistic form of retributivism (see Feinberg, “What, if Anything, Justifies Legal Punishment,” pp. 613-14), which form few retributivists would embrace.

¹⁵ The term “split-level,” as well as its contrast with “mere conjunction” and “integrated” accounts, which terms I also adopt, comes from Wood (2002).

¹⁶ See, e.g., Hart, *Punishment and Responsibility*, p. 3 (“[I]t is likely that in our inherited ways of talking or thinking about punishment there is some persistent drive towards an oversimplification of multiple issues which require separate consideration. To counter this drive what is most needed is *not* the simple admission that instead of a single value or aim (Deterrence, Retribution, Reform or any other) a plurality of different values and aims should be given as a conjunctive answer to some *single* question concerning the justification of punishment.”); Wood, p. 303. A much more approving view of the additive or conjunctive approach is expressed in John Gardner, “Crime: in Proportion and in Perspective,” in *Fundamentals of Sentencing Theory* eds. A. Ashworth & M. Wasik (1998), pp.32-33.

¹⁷ For the conclusion that integrated dualism is conceptually unobtainable see Wood, pp. 316-17. A scent of integrated dualism is apparent in Larry Alexander, “The Doomsday Machine: Proportionality, Punishment and Prevention,” *Monist* 63 (1980): 199-227.

This paper offers an integrated dualist theory of punishment.¹⁸ In doing so, it assumes that the road to a satisfactory account of punishment’s justifiability must start by examining with seriousness the common point of departure already noted – namely, that punishment stands in need of justification. Frequently, although not universally, theorists expend some effort trying to define what punishment is. Very rarely, however, do they attend explicitly to what it means for punishment (or anything else, for that matter) to be “in need of justification.”¹⁹ That is a most unfortunate oversight, for, as Hart recognized, “theories of punishment” is an unhappy term because such “theories” “are not theories in any normal sense. . . . On the contrary, . . . [they] are moral *claims* as to what justifies the practice of punishment.”²⁰ Yet more precisely, they are moral claims *in response* to the proposition that punishment stands in need of justification. If theories of punishment are thus situated *ab initio* within an argumentative dialectic, one might expect their persuasiveness to depend, in part, on how fully and satisfactorily they understand the proposition to which they aim to respond.

* * *

Human practices, like other phenomena, often encompass a core and a periphery. Accordingly, Part I distinguishes core from peripheral cases of punishment. Part II then seeks to

¹⁸ Although the account is dualist, the elements of the argument likely to prove most interesting and contentious arise in support of the role I assign retributivism. So the argument can also be viewed as a rehabilitation of retributivism, though as a necessarily partial one.

¹⁹ Kleinig, ch. 1, is a notable exception. As he put it, to say that punishment stands in need of justification “is not to say anything very helpful, without first saying something about the notion of justification to see what it is, precisely, that punishment is said to be in need of.” *Ibid.*, p. 1.

²⁰ Hart, *Punishment and Responsibility*, p. 72.

unearth the logic of justification. Such a logic must address two questions: what creates a need of justification, and how can such a need be met? With regard to the former, I argue that for some practice to stand in need of justification, considerations must first be identified that put the overall justifiability or permissibility of that practice in doubt. I term any such consideration a “demand basis.” With regard to the latter, I explain that justification can be supplied by showing either that the putative reason that constitutes a demand basis – i.e., that brings forth the need for justification – does not apply or that such a reason is outweighed by competing considerations. In Razian terms, justification can be supplied by cancelling conditions or by overriding reasons.

Together, Parts III and IV seek to justify punishment in light of Part II’s analysis of justification. Part III identifies the demand basis that underlies the proposition that punishment stands in need of justification – namely, that it inflicts suffering – and unpacks it into two discrete claims – that punishment is wrong because of the bad that it causes (pain, suffering, the deprivation of liberty), and that punishment is wrong because the intentional infliction of that bad infringes the defendant’s rights. Part IV then argues that core cases of punishment are justified by cancellation on retributivist grounds and that peripheral cases of punishment are justified by override on consequentialist grounds. More particularly, core cases of punishment are justified because the offender’s desert renders the fact that punishment inflicts suffering wholly inert as a reason against it – i.e., the offender’s suffering is not a bad and no right of his is infringed. In peripheral cases, by contrast, the offender’s suffering is a bad and therefore does count as a reason against the punishment. So even if, as I argue, punishment need not be rights-infringing even in peripheral cases, if peripheral cases of punishment can be justified against this demand

basis, it must be because the consequential benefits produced by such punishment have overriding weight.

As Part V explains in greater detail, the different modes of justification (cancellation and override) and the different instances of punishment (core and periphery) thus combine to make dualism plausible, perhaps compelling. Retributivism and consequentialism are integrated components of the response to the demand that punishment be justified on account of the suffering it inflicts. In summarizing the account, Part V also addresses anticipated objections.

I. PUNISHMENT

What is punishment? Or, to invoke the usual qualification, what is *criminal* or *legal* punishment?²¹

Philosophers of a generation or two ago paid considerable attention to the question, partly “because some arguments about its justification ha[d] been supposed to rest on how it is defined.”²² We no longer think that. As R.A. Duff has commented, the effort at definition “is doomed to futility if it is intended to produce a definition capturing all and only those practices that properly count as ‘punishment’, and that must rapidly become a normative discussion of how

²¹ Throughout, I am interested only in *criminal* punishment, or punishment inflicted by a state for the (claimed or supposed) violation of a criminal norm. (By criminal “norm,” I mean to capture cases, as in Nuremberg, in which a punitive response is imposed for wrongdoing not covered by previously enacted law. I mean to sidestep the debate over whether such moral norms are part of the law.) I am not concerned with, for example, “punishment” of children by parents, although what I say about criminal punishment may be relevant for other cases too.

²² Honderich, p. 6 n.4 (citing several of the best known attempts at definition).

punishment can be justified if it is to produce a useful account of what we *should* mean by ‘punishment’.”²³ Accordingly, with Duff, let us treat punishment as “something intended to be burdensome or painful, imposed on a (supposed) offender for a (supposed) offense by someone with (supposedly) the authority to do so.”²⁴

Although a loose definition of this sort is, I think, good enough, we needn’t treat without differentiation all cases within its ambit. To the contrary, the justificatory project is aided by distinguishing core (or central) cases from peripheral ones. The core case – symbolically represented as P_c – involves imposition of something painful or burdensome on an actual offender on account of his offense.²⁵ Furthermore, I will treat the core case as including only impositions that satisfy some form of proportionality constraint, whether of a retributivist or Benthamite variety.²⁶

Peripheral cases— P_p for short—involve the imposition of punishment on persons who are not wrongdoers, either because the supposed offense did not occur or because, although it did, the individual in question was not responsible for it. I mean “not responsible” broadly. I treat as peripheral those cases in which the person punished was not even a factual cause of the offense (as in cases of mistaken identity), or does not satisfy all the offense elements (as in cases of

²³ Duff, *Punishment, Communication, and Community*, p. xiv.

²⁴ *Ibid.*, p. xiv-xv. See also Lacey, pp. 11-12.

²⁵ See also Ten, p. 2; D.J. Galligan, “The Return to Retribution in Penal Theory,” in *Crime, Proof and Punishment* ed. C.F.H. Tapper (1981): 144-171, p.152 n.3.

²⁶ Although operationalizing a principle of proportionality—especially of a nonconsequentialist vein—presents well-known difficulties, for want of space I will assume without argument that the basic notion is coherent.

mistake that negate mens rea) or possesses a valid justification defense (as use of force in self-defense) or lacks moral responsibility (as in cases of insanity or infancy). I also include cases in which the imposed punishment violates whatever proportionality principle the theory in question purports to adhere to.

Peripheral cases all involve genuine adjudicative or legislative error. (It is this that places them at the margin of the concept as we have it.) In each, the state believes that the individual to be punished is morally responsible for an offense, and that the punishment to be imposed is commensurate with that responsibility. Were the state correct, we would have a core instance of punishment. That the state is incorrect creates a peripheral instance. Additionally, cases exist in which the state imposes on an individual who is not morally (or even causally) responsible for an offense, aware (or strongly suspecting) that the individual is not responsible. In this category include the much-discussed cases of scapegoating, and cases of plainer evil, as when one or more agents of the state's criminal justice apparatus, moved by animus or desire for personal gain, cause punishment to be inflicted on someone they believe innocent. I treat these as degenerate cases of punishment and will put them aside.²⁷

In short, then, let us treat punishment as the sum of core and peripheral cases.²⁸ The line

²⁷ Compare Lacey, p. 6, who conjoins peripheral and degenerate cases.

²⁸ This is not to engage in the “definitional stop” rightly criticized by Hart, *Punishment and Responsibility*, p. 5, for I am not contending that degenerate cases (call them P_d) fall outside the bounds of “punishment” by virtue of neutral or natural semantic criteria. My choice to eliminate degenerate cases from the scope of “punishment” is a stipulation designed to simplify exposition, on the assumption that most commentators are agreed that P_d is not justifiable. Therefore, allow “punishment” to stand for only those forms of punishment whose justification is in doubt and that I will try to establish – namely, $P_c + P_p$. But see *infra* note [84] (problematizing

between core and periphery likely could be drawn with greater precision. For present purposes a rough first take will be good enough. The remainder of this paper is premised on the idea that to separately justify P_c and P_p is to thereby justify punishment.

II. JUSTIFICATION

Punishment stands in need of justification. Call this proposition $J_n(P)$. I assume that this is true of core and peripheral cases. Thus: $J_n(P_c)$; and $J_n(P_p)$. In order to understand how these needs could be satisfied we must be clear about the nature of the need. What, we might wonder, does it *mean* for punishment to require justification? Or, more fundamentally, what does it mean for *anything* to be in need of justification?

A.

Justification can always be given. There is always some sense in offering the reasons that support one's conduct.²⁹ But, though they may always be offered, justifications cannot always be *demand*ed. Customarily, to claim that any actual or contemplated conduct, X , is "in need" of justification means that "ordinarily" or "presumptively" one ought not to X . It means, at the least, that there exists apparent or putative reason not to X .³⁰ In other words, our practice of

 this straightforward carving between peripheral and degenerate cases).

²⁹ Although justifications are as important for beliefs as for action, I will be interested only in how they function in practical reasoning, leaving wholly aside how, if at all, the analysis offered would differ in theoretical contexts.

³⁰ One might suppose that it means more than this. Perhaps to affirm that X is in need of justification means not merely that there is apparent reason not to X (or that we have reason to suspect not to X), but that we have reason not to X . I will consider this objection *infra* Section V.D. For now, I stick with the formulation in text. In doing so, I also reject the possibility that to

justification is a dialectical one in which the argumentative burden rests, in the first instance, on those who would deny justification.³¹

Or perhaps you believe, contrary to what I have supposed, that every action *is* always in need of justification. If so, let us identify and distinguish two concepts: Call the always-applicable need to justify a course of action a requirement for “general justification,” and the need to justify action in response to articulable doubt a call for “specific justification.” The form of justification particularly at issue in debates over criminal punishment is specific, not general. Were the case otherwise, it would be hard to explain why the literature on punishment starts with

demand punishment means less than I have supposed—in particular, that mere puzzlement about the propriety of X, short even of articulating a possible reason against X, is enough to generate a demand for justification.

³¹ As John Gardner has written: “In a loose sense, justification is always called for. That is just to say that actions, beliefs, etc. are always answerable to reason. One may always ask ‘why?’ But in the stricter and more important sense which concerns us here, justification is called for only when one also has some reason *not* to act, believe, etc. as one does. The unobjectionable, in other words, is in no need of justification.” John Gardner, “Justifications and Reasons,” in *Harm and Culpability*, eds. Simester & Smith (1997): 103-129, p. 107.

I fully agree with what I take to be Gardner’s core point, namely that although there is a sense in which action can always be justified, or not, that’s not the sense that concerns us. For one to say that “X is in need of justification” means something more than the “loose sense” conveys. But I’d add two caveats. First, to observe that the unobjectionable does not need justification might be thought to beg the question, which is whether given conduct *is* unobjectionable. After all, it could be that every action and belief is in need of justification precisely to establish that it is not objectionable. We might rather say either that the *seemingly* unobjectionable, or the actually *unobjected-to*, is in no need of justification.

Second, I would not go so far as contending that the “stricter” or “more important” sense of justification is implicated only when one does in fact have reason not to X. Rather, as the preceding footnote suggests, I would leave open the possibility that the question of justification can be raised even when, fully understood, there is *not* reason not to X, though one mistakenly believes or anticipates that there is.

Jn(P), whereas the literature on, say, chess does not start with the proposition Jn(C). We strive to justify punishment in the face of articulable suspicion that the task might not be doable.

It follows from the principle that only the seemingly objectionable must be justified (or from a focus on specific as opposed to general justification) that every claim that X is in need of justification stands, implicitly if not always explicitly, for a fuller claim that contains, as well, the basis for the demand. Joel Feinberg famously taught that a claim of personal desert necessarily contains not only identification of the treatment said to be deserved (punishment, six years at hard labor, an A-) but also the “desert basis”: the personal characteristic or conduct *in virtue of which* that treatment is deserved.³² Demands of justification likewise contain a “demand basis”: the consideration that gives rise to the demand. The statement that “X is in need of justification” is therefore incomplete. To make explicit what is implicit, we should say that “X is in need of justification, by virtue of y.” Let us represent such a demand for justification thus: Jn(X)(y).

Armed with this fuller understanding of what a demand for justification means, we can address how such a demand is satisfied. Let us consider first *how* justification can be supplied in the face of a demand. By what *means* can a demand for justification be satisfied? A second and related question concerns the magnitude of the justificatory burden. Precisely what, or *how much*, must be established by one upon whom a justificatory demand rests?

³² Joel Feinberg, “Justice and Personal Desert,” in *Doing and Deserving* (Princeton: Princeton University Press 1970), p.58.

B.

There are at least two ways to meet $J_n(X)(y)$, two ways to answer a properly crafted justification demand.

The first and more obvious way is by supplying *overriding* reason to X. Surely one ordinarily or presumptively ought not to knowingly cause the death of another human being. Knowingly causing the death of another human being, we might say, stands in need of justification. Just as surely, however, this need can sometimes be satisfied. This is well illustrated by paradigmatic examples of the justificatory criminal law defense termed “necessity.”

Consider, for example, the vaunted Trolley Problem. If the only way to save the lives of the five sleeping workers is to turn the trolley onto a spur where a single worker lies asleep, it is not the case that, on balance, one ought not to knowingly cause the death of that one worker by turning the trolley. That is to say, the fact, z, that turning the trolley is necessary to avoid death to a greater number justifies knowingly causing the death of the one. It does so by furnishing a reason to act in a way that would produce the individual’s death that is of greater weight than the reason not to so act. In familiar terms, fact z supplies an overriding reason to X.³³

But there is a second way in which some fact, z, can render it untrue that, on balance, one ought not to X. In Razian terms, z can constitute a cancelling condition. Consider Raz’s own

³³ More than one reader has pointed out to me, rightly, that fact z meets the demand for justification so long as it supplies reason of *at least equal* weight to the countervailing reason contained in the demand basis. For this reason, I acknowledge that it would be more precise to describe the first way of satisfying a demand for justification as the supplying of a *sufficient* or *adequate* reason, where sufficiency or adequacy is defined as a countervailing reason of equal or greater weight. Sacrificing a slight degree of precision for greater terminological familiarity, I will, however, continue to speak of *overriding* reasons, not of *sufficient* or *adequate* ones.

example.³⁴ Ordinarily or presumptively one ought to act in accord with one's promise, say a promise not to X. But if the promisee has freely released the promisor from her promissory obligation, the promise itself no longer constitutes a reason not to X. The release is said to cancel the reason not to X that was created by the fact of the promise. Or, one ordinarily has a reason not to punch another in the nose. But if A and B have voluntarily entered into a boxing match, the background reasons not to pummel each other are cancelled. As these examples suggest, whereas necessity supplies the paradigmatic example of an overriding reason, consent supplies the paradigmatic example of a cancelling condition.

The difference can be put this way: when an action X, prima facie wrongful, is justified by overriding reason, the demand basis, y, remains a reason to not X. Not so when X is justified by a cancelling condition. "A cancelled reason is not even a prima facie reason for action."³⁵

C.

One might suppose that if X stands in need of justification then a defender of X has not satisfied her justificatory obligation until she shows that X is right, or at least not wrong, *all things considered*. Is this correct? Does invocation of a demand for justification impose on a

³⁴ Joseph Raz, *Practical Reason and Norms* (Princeton: Princeton University Press, 2d ed. 1990), p.27.

³⁵ Michael S. Moore, "Authority, Law, and Razian Reasons," *Southern California Law Review* 62 (1989): 827-896, p. 846. As Raz noted: "We often feel that a reason which has been cancelled has been cancelled completely and is no longer a reason whereas a reason which 'merely' overridden is somehow still there, it somehow survives." Raz, *Practical Reason and Norms*, pp. 202-03. This is a signal difference between cancelled reasons on the one hand and outweighed and even excluded reasons on the other. See also John Gardner & Timothy Macklem, "Reasons," in *The Oxford Handbook of Jurisprudence and Philosophy of Law* eds. Jules Coleman & Scott Shapiro (Oxford 2002): 440-475, p. 464.

would-be justifier such a heavy (or broad) argumentative burden?

I think it does not. Rather, the principle that only the seemingly objectionable must be justified entails that one's justificatory burden is demarcated by the considerations that provoked a need for justification in the first place. That is, when justification is offered in response to a demand therefor, such justification is accomplished so long as the considerations that, at least implicitly, brought forth the need for justification are satisfactorily addressed, either by conceding that those particular considerations *are* reasons and arguing that they are overridden, or by establishing that they rest upon only *apparent* reasons that are cancelled. We might say that special justifications are "tailored," not all-things-considered. It follows, of course, that they are necessarily provisional. What appears to be justified might turn out not to be if the skeptic can provide new apparent reasons why the action is wrong. Justification for a practice is always partial, always particular to the considerations advanced to bring forth the demand, always responses to particular demand bases.

Consider an example. I'd expect you to agree that my having showered this morning does not stand in need of justification. If so, that is not because I am unable to give reasons aimed to show that, all things considered, my showering was right or morally permissible. But the fact that I could give reasons for having showered does not mean, except in Gardner's "loose sense," that such reasons are demanded. Justification is possible, but not required.

Suppose however that there's a drought. Now we might fairly say that my having showered stands in need of justification. That is true but, as I argued above, incomplete, for it omits the demand basis. More precisely, then: my having showered this morning requires

justification in virtue of the stronger-than-usual need to conserve water. Formally:

Jn(Shower)(drought), or Jn(S)(d).

Suppose further, however, that I had fallen into a patch of poison ivy. The only way to escape an uncomfortable rash was to wash promptly and thoroughly. Under these circumstances, I expect that my showering was justified despite the drought.

Is this conclusion premature? After all, there could exist other reasons why I should not have showered notwithstanding the need to cleanse myself of the allergen in poison ivy. Perhaps, for example, today is Tisha B'av, a Jewish day of mourning. If I had promised my father, an observant Jew, to observe the applicable religious proscriptions, and if one of them concerns washing, and if, under Jewish law, that proscription is binding even at the cost of an annoying but not life-threatening rash, then my having showered presumably was not justified all things considered. Or perhaps I knew that corrosion in my bathroom plumbing threatened leaking upon my downstairs neighbor. Under these conditions, I might not have been justified in showering in my own shower (as I did), though I would have been justified in showering at my health club.

Here is the question: Once we allow that my poison ivy encounter meets the objection that my showering was not justified by virtue of the drought, are we not entitled to suppose that my showering was justified *simpliciter* until I foreclose all possible considerations that could *conceivably* render my showering wrong? Must I now establish, among innumerable other things, that today is not Tisha B'av (or that I have made no promise to observe Tisha B'av), and that my pipes are not leaking into my neighbor's apartment?

I should think not, for a contrary view would undermine the distinction between (what I

have called) specific and general justification. A need for general justification arises if a practice must be shown to be right, or permissible, in advance of any reason for doubt – absent, that is, any demand basis. Special justification is given in response to a demand basis. Because general justification does not arise against, or in response to, a reason or putative reason against the practice, I suppose that it must aim to meet all possible objections. A general justification must purport to be all-things-considered, for it is not clear how it could be partial or limited without thereby implicitly conceding that it is a special justification not a general one.

But our justificatory practice, I have claimed, overwhelmingly involves special justification not general. If we do not believe that we labor under an always-present demand to justify all our actions and beliefs, it is hard to see why identifying *some* reason to believe that X is wrong would thereby create a burden to establish that every conceivable but uninvoked reason against X is invalid, either overridden or cancelled. In fact, we have affirmative reason to believe otherwise, for one of the strongest arguments against a practice of demanding general justification is precisely that the all-things-considered justification at which general justification must aim is far too onerous. Put another way, a critical reason to favor a justificatory practice in which justification is demanded only in response to particular demand bases, and not as a background or default condition, is that this is the only sort of justificatory practice that can accommodate the justifications themselves being tailored as opposed to all-things-considered.

There is nothing odd about concluding that to demand justification is to rely on one or more demand bases, and that justification is thus (provisionally) supplied so long as the particular demand bases at work are defeated. It is to say only that justification, like a range of

concepts, is relational.³⁶ Here, it relates to the demand basis. When speaking precisely, then, we ought not to say, simply, that “X is justified” unless we intend to claim all-things-considered justification. Ordinarily, we are engaged in special justification and should accordingly say only that X is justified against, or notwithstanding, y. We can represent this statement with the notation $J(X)(y)$. That is, $J(X)(y)$ denotes a satisfactorily tailored response to $J_n(X)(y)$.

* * *

We have identified three features concerning the making of, and responding to, demands for justification. First, to demand that an action or practice be justified requires invocation of, or implicit reliance on, a reason to believe that the action or practice is wrongful. A demand for justification depends on a “demand basis.” Second, we meet the demand for justification either by overriding the demand basis or by cancelling it. Third, if the demand basis is defeated, the demand for justification retains no additional force; it imposes no further obligation on the

³⁶ A particularly well known example is Dworkin’s discussion of discretion:

The concept of discretion is at home in only one sort of context; when someone is in general charged with making decisions subject to standards set by a particular authority. . . Discretion . . . is therefore a relative concept. It always makes sense to ask, ‘Discretion under which standard?’ or ‘Discretion as to which authority?’ Generally the context will make the answer to this plain, but in some cases the official may have discretion from one stand-point though not from another.

Ronald Dworkin, *Taking Rights Seriously* (Harvard Univ. Press 1977), p.31. Much the same can be said about justification. It is at home in the context of someone (perhaps the would-be justifier herself) having questioned the permissibility of an action. And it therefore always makes sense to ask “Justification against which demands?” for the action might be justified against one demand though not against another. See also Kleinig, p. 3 (“demands for justification generally relate to *particular features* of a situation rather than the situation *as a whole*, and . . . it is only in the latter case that we can confidently speak of having justified a piece of behaviour”).

proponent of a practice. This is not to say the practice is thereby rendered justified, or even justifiable, all things considered. Maybe it is, maybe not. The more modest (yet significant) upshot is only that the practice no longer stands (embarrassedly) “in need of justification.” Indeed, if we generally presume actions to be permissible absent at least an argument fragment to the contrary (as I submit we do), the argumentative burden shifts back to the practice’s doubters.

III. THE DEMAND BASIS: PUNISHMENT INFLICTS SUFFERING

Why is punishment in need of justification? By virtue of what is Jn(P) true? This would not seem mysterious. As Hart put it, agreeing with Benn, it is “the deliberate imposition of suffering which is the feature needing justification.”³⁷ Indeed, few claims in the philosophy of criminal law are more widely accepted. Theorists routinely observe not merely that “punishment stands in need of justification,” but that such justification is needed precisely “because it involves the infliction of pain or other form of unpleasant treatment.”³⁸ As a first pass, then, let us say that

³⁷ Hart, *Punishment and Responsibility*, p. 2 n.3.

³⁸ Finkelstein, quoted *supra* note [2]. A tiny but representative sampling of claims along these lines includes Igor Primoratz, *Justifying Legal Punishment* (1989), p.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”); Richard Wasserstrom, “Why Punish the Guilty?” in *Princeton University Magazine* 20 (1964): 14-19, reprinted in *Philosophical Perspectives on Punishment*, ed. Gertrude Ezorsky (Albany: SUNY Press 1972): 328-41, p. 337 (“Punishment is an evil, an unpleasantness; it requires that someone suffer. Its infliction demands justification.”); Gertrude Ezorsky, “The Ethics of Punishment,” in *Philosophical Perspectives on Punishment*: xi-xxvii, p. xi (endorsing McTaggart’s observation that “punishment is pain and to inflict pain on any person obviously needs justification”); Lacey, p. 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

the demand basis supporting Jn(P) is that it inflicts suffering upon the individual punished.

This is assuredly not to claim that the fact that punishment causes suffering to the individual punished is the only possible, or even actual, mark against it. In addition, punishment causes pain and hardship to individuals who care about the individual in question or who depend upon him. And it costs money that could be spent by the government on other projects or left in the hands of citizens. Punishment might also brutalize or degrade the society that authorizes it, and the individuals who inflict it.³⁹ The list could be expanded. But no doubt several of the considerations that seem to weigh against punishment count against scores of other governmental activities, from imposing taxes to building roads. Surely nobody could fail to notice that the claim that punishment is in need of justification is couched in terms that suggest a particular status, character or urgency. We do not say (merely) that “punishment, like every state practice, demands justification.” Rather, as Richard Burgh 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. It is in virtue of this that the institution requires justification *in a way that many other political institutions do not.*”⁴⁰ It is the fact that

consequences . . . and doing so irrespective of the will or consent of the person being punished.”); Crocker, p. 1063 (“From a moral point of view there is something initially puzzling about the very existence of punishment. Punishment is, after all, the intentional infliction of suffering, or at least some deprivation, upon a human being.”).

³⁹ These sorts of reasons against punishment have been emphasized in, for example, Husak, *Why Punish the Deserving?*, and Honderich, p. 7 (“The problem of punishment arises mainly *but not only* for the reason that the practice involves what traditionally has been called suffering”) (emphasis added).

⁴⁰ R.W. Burgh, “Do the Guilty Deserve Punishment?” *Journal of Philosophy* 79 (1982): 193-210, p. 193.

punishment involves the intentional infliction of suffering that *particularly* demands justification. The claim that, I observed at the outset, motivates almost the entire philosophical literature on punishment can therefore be reformulated with somewhat greater precision by recasting Jn(P) as Jn(P)(inflicts suffering), read as “Punishment stands in need of justification on account of the fact that it inflicts suffering.”

Yet still more precision is required, however, for we should wish to know what about the fact that punishment inflicts suffering constitutes a reason against it. The obvious utilitarian answer – and the answer likely first to appeal to a welfarist of any stripe – is that the state of affairs in which an individual suffers is a bad state, one which we all have reason not to bring about. This seems right as far as it goes, but it does not go far enough.

Yes, punishment causes suffering. But it does so in a particular way – namely, by “inflicting” it, or (if this is not redundant), by “intentionally,” “purposefully,” or “deliberately,” inflicting it. The theorist’s frequent recourse to terms of this sort would seem to suggest that they capture something of importance.⁴¹ We might be tempted to say, then, that what makes the fact that punishment intentionally inflicts suffering a reason against it is not that suffering is an intrinsically bad state of affairs, but that the intentional infliction of suffering is an intrinsically

⁴¹ Consider also that the suffering that supports Jn(P)(inflicts suffering) is particularly the suffering experienced by the person punished, not the suffering experienced by that person’s friends, lovers, dependents, on account of that person’s suffering. And this is true even if the suffering experienced by those others exceeds in magnitude the suffering experienced by the punished person himself.

wrongful action. And it is wrongful, some say, because it violates an individual's rights.⁴²

But to substitute this explanation for the first would be too quick. First, as I have been at pains to emphasize, the proposition that punishment stands in need of justification is common ground among all theorists. We should therefore strive to construe it in a way that makes sensible its endorsement by persons approaching from as wide a range of normative perspectives as possible. But to construe Jn(P)(inflicts suffering) in this way might make it unacceptable to some, for the extent to which consequentialists of various flavors can embrace rights claims remains contested. Furthermore, even those fully comfortable with rights discourse should wish not to lose sight of the straightforward consequentialist objection. After all, while proponents of the doctrine of double effect believe that there is a significant moral distinction between intentionally inflicting suffering and knowingly causing it, not even they think that the latter is of no moral import. Any practice, Q, that predictably causes suffering can properly, for that reason alone, be said to need justification.

For these reasons, we do best to view the statement represented by Jn(P)(inflicts suffering) – that “punishment stands in need of justification on account of the fact that it inflicts suffering” – as capturing two distinct but related demand bases. Punishment stands in need of

⁴² Naturally, the intentional infliction of suffering could be wrongful for a variety of reasons other than the badness of the suffering that it causes. What could be peculiarly bad about the intentional infliction of suffering, for example, is that it brutalizes or debases those who authorize and impose the sanction in a way that (merely) knowingly causing the bad of suffering does not. I put this aside not because I think the view wrong but because I think it represents, not an elucidation of the demand basis captured by Jn(P)(inflicts suffering), but rather a distinct demand basis, just as would the fact (plainly true) that punishment is expensive. And I think this is so because I construe Jn(P)(inflicts suffering) to be grounded in concern for the interests of the persons punished.

justification both on account of the fact that it causes the punished person to suffer, and on account of the supposed fact that, by intentionally inflicting suffering, it infringes an individual's rights.⁴³ For brevity of exposition these discrete facets or construals of the claim $J_n(P)$ (inflicts suffering) can be formalized, respectively, as $J_n(P)$ (suffering) and $J_n(P)$ (infliction).

That punishment inflicts suffering is true in core and peripheral cases alike. With regard to the demand basis represented by the fact that punishment inflicts suffering, that is, core and peripheral cases are indistinguishable. Moreover, both facets or construals of this demand basis seem to apply in both cases. So, by starting with the proposition that punishment stands in need of justification on account of the fact that it inflicts suffering, and then (a) dividing punishment into core and peripheral cases, and (b) distinguishing two reasons why the infliction of suffering generates a demand for justification, we are left with four distinct demands for justification. Symbolically, $J_n(P)$ (inflicts suffering) breeds, or is shorthand for: (1) $J_n(P_c)$ (suffering); (2) $J_n(P_c)$ (infliction); (3) $J_n(P_p)$ (suffering); and (4) $J_n(P_p)$ (infliction). Each of these four claims requires a response if punishment is to be justified against the fact that it inflicts suffering.

IV. JUSTIFYING PUNISHMENT AGAINST THE FACT THAT IT INFLICTS SUFFERING

A.

Let us start with $J_n(P_c)$ (suffering), read as "core cases of punishment stand in need of

⁴³ As we will see, these demand bases, though related, are fairly understood as logically independent, such that the intentional infliction of suffering is bad or wrong even when the suffering thus inflicted is not a bad (and vice versa). Put another way, the badness or wrongness of inflicting suffering is not to be understood as merely an aggravated form of the badness or wrongness of causing suffering.

justification on account of the fact that they cause the punished person to suffer.” It is close to an analytic truth (problematized only by the fact of masochism) that suffering is experienced as a bad. For hedonists and most other welfarists, accordingly, the experience of suffering is an intrinsic bad. Hence the common supposition that the imposition of suffering by means of criminal punishment can be justified only by the presence of overriding reason.⁴⁴ But that suffering is *always* an intrinsic bad seems to be what most retributivists deny. To be sure, precisely what retributivism is or contends is much debated. If not an essentially contested concept, it is, at minimum, radically controverted. Still, if contemporary retributivism contains a core claim, it must be that offenders deserve to suffer.⁴⁵

⁴⁴ Unfortunately, John Kleinig – the theorist who understood most clearly that those who would speak intelligently about justifications for punishment need first seek to understand the nature of justifications – seemed to assume precisely this. “Impositions require justification since they constitute breaches of a basic moral right to be free from such things,” he wrote. But this “does not mean that it cannot be overridden in any circumstances.” Kleinig, p. 65. That he thought the justification of punishment must proceed by override is not at all surprising, however, given his mistaken view (mistaken because seeming to overlook that justification can be accomplished by cancellation) that “[d]emands for justification are demands for reasons.” Ibid., p. 2. See also, e.g., Lacey, *State Punishment*, p. 14 (“[O]ur primary task is to establish whether there are reasons for having institutions of punishment and for inflicting punishment in particular instances, which outweigh the reasons which militate against them.”); Finkelstein, p. 358 (“A criterion of adequacy for any theory of punishment . . . is that it provides an account of when punishment overcomes the presumption against the moral legitimacy of the acts it involves.”). I take the words “outweigh” and “overcomes” to likewise imply that justification for punishment must proceed by means of overriding reason.

⁴⁵ Bagaric & Amareskara, p. 127 (“A vast array of theories of punishment have been advanced which are classified as retributive. As a result of the diversity of these theories, it has proven remarkably difficult to isolate a distinctive feature of theories which carry the tag. Nonetheless, all retributive theories assert that offenders deserve to suffer”) (footnotes omitted); Christopher, “Deterring Retributivism,” p. 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.”); Kent Greenawalt, “Punishment,”

Criminal Law & Criminology 74 (1981): 343-362, p. 347 (defining retributivism as the theory that “punishment is justified because people deserve it”); Primoratz, p. 12; Ten, p. 5 (For a retributivist, “[t]he offender’s desert . . . is what justifies punishment.”); *ibid.*, p. 46 (“Contemporary retributivists treat the notion of desert as central to the retributivist theory, punishment being justified in terms of the desert of the offender.”); Michael S. Moore, *Placing Blame* (1997), p.88; Alexander, “The Doomsday Machine,” p. 199; (“Retributivists argue that punishment must be justified by the ill-desert of the one punished.”); Hugo Adam Bedau, “Retribution and the Theory of Punishment,” *The Journal of Philosophy* 75 (1978): 601-20, p. 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.”). As Cottingham, “Varieties of Retribution,” makes clear, desert has not always been so central to views historically classified as retributivist. But the centrality of desert to contemporary retributivism is hard to gainsay.

As others have noted, many theorists describe what is deserved as *punishment*, not as *suffering*. See, e.g., Douglas N. Husak, “Retribution in Criminal Theory,” *San Diego Law Review* 37 (2000): 959-986, pp. 972-973 (arguing—rightly, in my view—for the latter over the former). Although this is not the place to fully examine the relative merits and demerits of each formulation, a few words seem necessary. As a justification for punishment, the claim that wrongdoers deserve *to suffer* runs into the objection that, in principle, the state would be required to adopt what Gertrude Ezorsky called the “whole life” view of criminal desert: The suffering imposed by means of criminal punishment would have to be reduced or wholly eliminated if the individual had already experienced as much suffering in his life as his wrongdoing merited. See Ezorsky, pp. xxii-xxvii; see also W.D. Ross, *The Right and The Good* (Oxford: Clarendon Press, 1930), pp. 58-59 (anticipating this objection). This unwelcome conclusion might seem to be avoided if what a wrongdoer deserves is *to be punished*. If suffering and punishment are distinct concepts, and if what the wrongdoer deserves is the latter, then no amount of the former can affect the propriety of retributive punishment. Furthermore, if wrongdoers deserve not merely punishment (whatever that is, exactly), but *state* punishment, then the case for state action is bolstered to at least some extent because failure to punish renders it logically impossible for this desert to be realized. On the other hand, it might reasonably be thought that the plausibility of the desert claim itself diminishes as the consequence said to be deserved becomes increasingly specific. The very notion of desert is mysterious and, as we will see, heavily dependent on intuitive judgments of intrinsic plausibility. To many, that wrongdoers deserve to experience suffering simply seems truer than that they deserve particular treatment at the hands of a contingent entity, the state.

In this paper I assume, without further defense, a somewhat middle position that wrongdoers deserve to suffer on account of their blameworthy wrongdoing. This construal of what wrongdoers deserve escapes the whole-life view of suffering partly but not fully. Pre-

Admittedly, this core claim is beset by ambiguity, such as whether deserved suffering is comparative or non-comparative, and responsive to legal or moral wrongdoing. Because I conceive it the most plausible version, I will assume that desert, in the most promising retributivist account, is noncomparative and moralistic. In short, I assume something like the following claim is true: A person who unjustifiably and inexcusably causes or risks harm to others or to significant social interests deserves to suffer for that choice, and he deserves to suffer in proportion to the extent to which his regard or concern for others falls short of the regard or concern properly demanded of him.⁴⁶ Call this the “desert claim.” It is often thought equivalent to the claim that for offenders to suffer (on account of, and in proportion to, their blameworthy

offense suffering is strictly irrelevant to the offender’s liability to punishment (except insofar as such suffering could, perhaps in cases of extreme childhood abuse, mitigate the individual’s blameworthiness). But if the state aims to ensure that punishment does not exceed what is deserved, in principle it must reduce the amount of suffering it would otherwise inflict by an amount equal to any non-penal suffering the offender experienced on account of the wrongdoing. This “on account of” formulation is intended to signal (although I hold to this view tentatively) that non-penal suffering is relevant to the calculus only when it is caused by features of the actor’s conduct that render it wrongful—vengeance inflicted by, or in the name of, his victim, for example, but not (to take an illustration offered by Scott Shapiro) a hernia incurred by the actor’s attempt to abscond with heavy goods.

⁴⁶ On this view, persons can deserve to suffer for moral wrongdoing that has not been criminalized in the individual’s jurisdiction. However, it does not at all follow, as critics frequently charge, that retributivists who adopt this view necessarily license the state in punishing such individuals. I don’t understand why retributivists are supposed to be unable to appreciate the value of the independent principle of legality. Nor does this view leave the retributivist unable to justify punishment in at least some cases of *mala prohibita*. Though the issue is obviously too complex to permit adequate treatment in a footnote, it is enough to observe that the law sometimes creates social interests deserving of respect that did not exist absent the law’s intervention. Knowing disregard of the law is thus sometimes, but not always, blameworthy in itself.

wrongdoing) is an intrinsic good.⁴⁷ (For ease of exposition, I will frequently omit the parenthetical qualifications in the rest of this paper, saying things like “wrongdoers deserve to suffer” and “the suffering of wrongdoers is an intrinsic good.” In either case, though, the qualifications that punishment be on account of and in proportion to blameworthy wrongdoing should be assumed.) On the sensible assumption that the suffering of a given individual cannot be, at the same time, both an intrinsic good and an intrinsic bad,⁴⁸ it follows that if offenders deserve to suffer, then the state of affairs in which they experience the suffering they deserve is not a bad state, and therefore one we lack reason not to bring about.⁴⁹

Such is the nutshell version of the retributivist argument, by cancellation, for

⁴⁷ E.g., Lawrence H. Davis, “They Deserve to Suffer,” *Analysis* 32 (1972): 136-40, p.136; Michael S. Moore, “Justifying Retributivism,” *Israel Law Review* 27 (1993): 15-49, pp. 19-20; Thomas Hurka, “The Common Structure of Virtue and Desert,” *Ethics* 112 (2001): 6-31. This “on account of” linkage captures the idea, noted by Feinberg (“What, if Anything, Justifies Legal Punishment,” pp. 614-15), that what is “good in itself” might involve context and relationship.

⁴⁸ In correspondence, John Deigh has objected that the proposition (1) that the state of affairs in which A experiences deserved suffering is good does not entail (2) that the state of affairs in which A experiences the suffering he deserves is not bad. The state of affairs can be good in one respect – namely, that A experiences what he deserves – and bad in another respect – namely, that A experiences suffering. To be clear, then, I am assuming that proponents of the desert claim deny that the experience of suffering is ever an intrinsic bad all by itself; the goodness or badness of the state of affairs in which one suffers is always contextual or relational.

⁴⁹ Turning Hart’s well-known objection to full-blooded utilitarians – that when we punish we should do so with the sense of sacrificing important principles or interests – against retributivists, Russell Christopher charges that “retributivism has no sense that punishment involves any choosing of the lesser evil or involves any sacrifice at all.” Christopher, “Deterring Retributivism,” p. 879. I agree, but with two caveats. First, this is true only for P_c , not for P_p . Second, even with respect to P_c , retributivists recognize that *some* sacrifice is involved. What they deny is that the inflicting of suffering on the wrongdoer is itself a sacrifice or evil. All other social costs, including but not limited to those I have already acknowledged, remain real sacrifices even for the retributivist.

J(P)(suffering). Historically, however, the desert claim has faced fierce criticism. According to anti-retributivists, the proposition that wrongdoers deserve to suffer on account of their blameworthy wrongdoing is mere *ipse dixit*. But this is both false and nearly beside the point.

The charge is false because retributivists do not simply *assert* that the suffering of wrongdoers is an intrinsic good. To start, they rely on actual incidents and thought experiments designed to evoke the case-specific substantive moral judgments that some particular wrongdoer deserves to suffer. John Kleinig, for example, asks us to imagine whether, stripped of consequential considerations, it is a better world in which Nazi war criminals live happily ever after or suffer on account of their atrocities.⁵⁰ Elaborating on similar examples of egregiously blameworthy wrongdoing, retributivists like Michael Moore then argue that these specific judgments or intuitions can be made most coherent by endorsing the general principle that wrongdoers deserve to suffer.⁵¹

⁵⁰ Kleinig, p. 67.

⁵¹ Michael S. Moore, "The Moral Worth of Retribution," in *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, ed. Ferdinand Schoeman (Cambridge: Cambridge University Press, 1987): 179-219. Because third-person emotions of resentment are often thought unreliable guides to correct moral judgments, Moore also employs thought experiments involving first-person wrongdoing so as to generate the presumably virtuous emotion of guilt. Although David Dolinko objects that relying on cases of especial savagery unfairly "stacks the deck," David Dolinko, "Some Thoughts about Retributivism," *Ethics* 101 (1991): 537-559, p. 557 & n.75, I think that any such objection reaches only the claim that we have a duty to punish (which claim I do not rely upon), not the more modest claim that blameworthy wrongdoers deserve to suffer on account of, and in proportion to, their blameworthiness. I suspect that the correct moralistic retributivist response to, for example, the "pedestrian who violates a state traffic law by crossing a street in the middle of the block – at 4:00 A.M., with no automobiles anywhere in the vicinity," *ibid.*, p. 556, is simply to deny that she deserves to suffer or, accordingly, to be punished. In cases of very modest blameworthiness (imagine, for example, our pedestrian darting out into traffic at 4:00 P.M) a correspondingly

More significantly, the anti-retributivist charge courts irrelevance because it wrongly presupposes that we possess resources for resolving questions of this sort that retributivists have failed to employ. As Bentham himself acknowledged, the intrinsic goodness of pleasure and the intrinsic badness of pain are matters to be assumed, not to be proven.⁵² To put the point more sharply, then, this particular anti-retributivist charge risks hypocrisy. Consider, for example, Herbert Fingarette, who objected in one breath to the retributivist reliance on unanalyzed intuition,⁵³ while, in the next, declaring it “self-evident” that “[r]etributive punishment is in itself an evil.”⁵⁴

The fact is that intrinsic goods just are not well suited to full resolution by rational

small degree of suffering (we might even prefer to term it an “unpleasantness”) could well be deserved, such as may be produced by a small fine. But even if this is so, the amount of deserved suffering will be so small that the example cannot be relied upon to generate strong intuitions. So theorists like Moore can hardly be faulted for employing cases of extreme wrongdoing in order to generate acceptance of the principle, which principle itself covers the inframarginal cases too.

A separate point, provoked by the fact that Moore himself clothes his retributivism in moral realism. In contrast, the argument that I am presenting for J(P_c)(suffering) does not depend upon any particular metaethics. Given the dialectical structure of justification, a justification for punishment need have no “greater,” more secure, or more objective, metaethical status than does the demand basis to which it responds. It follows as well that proponents of the dualist view I sketch here need not endorse metaphysical libertarianism.

⁵² Jeremy Bentham, *The Principles of Morals and Legislation* (1781), p. 4.

⁵³ Herbert Fingarette, “Punishment and Suffering,” *Proceedings and Addresses of the American Philosophical Association* 50 (1977): 499-525, p. 503.

⁵⁴ *Ibid.*, p. 513.

argumentation.⁵⁵ Retributivists believe that such suffering is an intrinsic good. Their critics – or some of them – believe otherwise. For them, suffering is always an intrinsic bad, making the avoidance of suffering an intrinsic good. In my view, we lack argumentative resources for resolving disputes of this kind, which is why the list of plausible candidates for intrinsic goodness is so long⁵⁶ and why appeal to casuistic judgments is unavoidable.⁵⁷

Thirty-odd years ago, Lawrence Davis succinctly adjudged the proposition that there is intrinsic value in wrongdoers' suffering "very likely" "on the grounds (a) that there is no convincing argument against it, and (b) that inclination to believe it seems very widespread among the people whose moral intuitions constitute the main data we have for settling questions of value."⁵⁸ Today's anti-retributivists are no closer to mounting a convincing counter-argument, while the ranks of those expressing such a belief have swelled. We therefore seem entitled to

⁵⁵ At least this is true, I believe, of intrinsic goods supposed to be ultimate. As Les Green helpfully reminded me, one can rationally justify an intrinsic non-ultimate good by showing how it is a constituent component of another intrinsic good (rather than as a means to its attainment).

⁵⁶ See, e.g., William Frankena, *Ethics* (2d ed. 1973), pp. 87-88.

⁵⁷ That such an appeal is unavoidable does not mean that we can do nothing more than argue about brute intuitions. Think, for example, of Thomas Hurka's elegant demonstration of the respect in which desert and virtue share a formal or mathematical structure. Hurka, "The Common Structure of Virtue and Desert." Still, even Hurka cautions against exaggerating the significance of his modeling, acknowledging that if we accept desert "it will be primarily on the basis of substantive judgments" about the superiority, as against an otherwise identical world, of a world in which the vicious suffer. *Ibid.*, p. 31.

⁵⁸ Davis, p. 139. See also *ibid.*, p. 140 ("Until someone provides better reason for picking and choosing among our apparently deeply felt moral convictions, the case for [the proposition that there is some intrinsic value in the suffering of the guilty] seems as solid as the case for any number of principles enjoying greater popularity at this date.").

conclude that $J(P_c)$ (suffering), on account of the fact that wrongdoers deserve to suffer.⁵⁹

B.

Even if contemporary anti-retributivists won't quite concede the desert claim, many do seem to concede that we are entitled not to affirm its negation; we are permitted to remain agnostic. Instead of continuing to tussle over whether wrongdoers deserve to suffer, then, they have tended increasingly to direct their critical fire elsewhere. In particular, they have begun more strenuously to emphasize that the fact that wrongdoers deserve to suffer, even if true, is not sufficient to justify punishment. Indeed, one prominent anti-retributivist, David Dolinko, deems

⁵⁹ Or are we? After all, strictly speaking, it is not the fact (if true) that an offender's suffering is an intrinsic good that cancels the fact of the offender's suffering from being a reason against punishment. What cancels that putative reason, hence supports $J(P_c)$ (suffering) is the somewhat more modest fact (if true) that an offender's suffering is not an intrinsic bad. And these are not logically equivalent. Even if "if offenders deserve to suffer then their suffering is not an intrinsic bad," is true, the converse is false. An offender's suffering could fail to be an intrinsic bad without it being the case that it is an intrinsic good – without it being the case, that is, that the offender *deserves* to suffer. Because the route from "an offender's suffering is an intrinsic bad" to "an offender's suffering is an intrinsic good" travels through "an offender's suffering is not an intrinsic bad," and because it is that intermediate proposition that provides cancellation, it might seem to follow that an offender's claimed desert plays no role in the justification of P_c against the suffering demand basis, in which event the structure of the argument is not retributivist. The assumption that offenders *deserve* to suffer might seem gratuitous.

To this charge I am disposed to plead guilty with an explanation. As a logical matter, it *is* the fact that an offender's suffering is not an intrinsic bad – and not the fact that it is an intrinsic good – that provides justification by cancellation. But it may well be the fact that an offender's suffering is an intrinsic good that furnishes the reason why his suffering is not an intrinsic bad. In other words, the spatial metaphor just relied upon might be unhelpful. It may not be that we reason, first, that an offender's suffering is not bad and only thereafter that such suffering is, indeed, good. It might be that the goodness of the offender's suffering is necessary to render that suffering not bad. In short, if suffering in a given case is not to be thought bad, the badness might have to be displaced, not merely denied. And if that's so, then the desert claim is necessary.

this “the principal objection” to retributivism.⁶⁰

This objection can be taken to mean – and, in the hands of some critics, is intended to mean⁶¹ – only that the fact that punishment brings about the intrinsic good of deserved suffering cannot establish that it is justified all things considered. This is surely true, but not – if what I have said so far about the structure of debates over the justification of punishment is correct – particularly meaningful. Retributivists take themselves to be offering a tailored justification for punishment – tailored, I have said, to the demand basis that punishment inflicts suffering. Of course, it may be that the ambition to offer only a tailored justification is misguided. But unless and until retributivists become persuaded that they are engaged in the wrong sort of enterprise, the proposition that the desert claim cannot establish that it is permissible all things considered for the state to inflict the deserved punishment looks more like an observation with which they are permitted to agree than an objection to which they need respond.

The contention that the desert claim is not sufficient to justify the infliction of punishment by the state holds more interest as an objection to the retributivist case if it is meant to invoke reasons for the insufficiency that are themselves represented by the demand basis that punishment inflicts suffering. If understood in this way, the objection is not a criticism of retributivists for seeking to offer only a tailored justification, but rather a charge that the desert claim does not accomplish even what the retributivists have sought. Not surprisingly, this is

⁶⁰ Dolinko, “Some Thoughts about Retributivism,” p. 558. Several other theorists who have endorsed this objection are cited in Christopher, “Deterring Retributivism,” p. 861 n. 85.

⁶¹ E.g., Husak (1992).

precisely how some proponents of this objection – Dolinko prominent among them – intend it. The desert claim is insufficient to justify punishment, they say, because the intentional infliction of suffering violates the rights of the person punished.⁶² And this is just to assert what I have termed Jn(P)(infliction). Having divided punishment into core and peripheral cases, we are therefore now poised to address Jn(P_c)(infliction).

1.

Or, almost poised. Before attempting to respond to this argument, we should get a clearer handle on just how it runs, for the simple contention that the intentional infliction of suffering violates the rights of the accused is too bald. To see why, consider the familiar distinction between infringing a right and violating it.⁶³ We say that a right is infringed when it is trampled upon, when the interests that give rise to it are harmed (and when the harm is traceable, in particular ways, to human agency). The infringement therefore gives rise to a demand for justification. If justification is satisfactorily provided, however, the right is not violated. A violation is a wrongful or unjustified infringement.

If we adopt this vocabulary – a vocabulary that does not commit us to any particular view about the type or weight of considerations that are required to justify the infringement of a right – we would not then say that the fact that wrongdoers deserve to suffer cannot justify punishment because the intentional infliction of suffering violates person’s rights (call this “the violation

⁶² See, e.g., Burgh, p. 198; Dolinko, “Some Thoughts about Retributivism,” pp. 543, 544, 558; Dolinko, “Three Mistakes of Retributivism,” pp. 1629-30.

⁶³ Judith Jarvis Thomson, *The Realm of Rights* (Cambridge: Harvard University Press, 1990), e.g., ch. 4.

claim”). This would be to state a conclusion that would be unacceptable to anybody – whether retributivist or consequentialist – who believes that punishment can be justified, which is to say that it would be unacceptable to just about everybody.⁶⁴ Put another way, the violation claim can be maintained by punishment skeptics but not by (mere) anti-retributivists. We should say, instead, that the fact that wrongdoers deserve to suffer (on account of and in proportion to their blameworthy wrongdoing) cannot itself justify punishment (against the fact that it inflicts suffering) because the intentional infliction of suffering infringes individual rights, or because people have a right against the intentional infliction of suffering even when such suffering is deserved.

Those wielding this claim think it congenial to consequentialists (as the violation claim would not be) because consequentialists know how to prevent an infringement from becoming a violation – namely, by supplying overriding reasons for engaging in the practice.⁶⁵ Although punishment infringes the right against the intentional infliction of suffering, it can produce so much good – most notably, by deterring antisocial conduct – as to prevent the infringement from becoming a violation. But the critics think that retributivists are unable to make a similarly plausible move. Even if the suffering of wrongdoers is an intrinsic good (which the critics might concede to be plausible), and even if the state has a legitimate interest in endeavoring to realize

⁶⁴ See, e.g., Simmons et al., “Introduction,” p. vii (observing that “most of us are confident that something like our current practice of punishment *can* be justified”). But see, e.g., Deirdre Golash, *The Case Against Punishment: Retribution, Crime Prevention, and the Law* (1995).

⁶⁵ Of course, this is true only for those consequentialists who do not find rights talk meaningless.

this particular good (which they might deem less plausible),⁶⁶ the claim that pursuit of this particular good is so important as to justify the infringement of rights is not plausible at all. After all, it is a mistake to assume that just because a good is intrinsic, it must be important as well. As Dolinko observed, even if “the pleasure experienced in orgasm is intrinsically good, . . . no one is likely to suppose we have even a *prima facie* reason to set up an official institution to provide people with orgasms.”⁶⁷ So much more fanciful is the idea that the effort to supply orgasms would justify the state’s infringement of individual rights. By the same token, the fact that wrongdoers deserve to suffer, or that their suffering is an intrinsic good, does not furnish a sufficiently compelling reason to warrant infringing the individual right against the intentional infliction of suffering.

To this important argument, retributivists would seem to have only two possible avenues of reply: to deny that the infliction of suffering in core cases of punishment infringes any right

⁶⁶ That this is not a legitimate state interest has been one recurrent theme in the literature. See, e.g., Jeffrie G. Murphy, “Retribution and the State’s Interest in Punishment,” *Nomos XXVII: Criminal Justice* eds. J.R. Pennock & J. Chapman (1985): 156-64. I cannot here give the argument the attention it merits, and must therefore content myself just to observe that there exist many plausible political theories that would deem the pursuit of any good *prima facie* legitimate. At the same time, I acknowledge that many political theories are not consistent with retributivist justifications for punishment. Nothing in this paper should be construed, then, either to claim that integrated dualism satisfactorily justifies punishment across the range of possible and actual theories of the role of the state, or to deny, as is increasingly being argued, see Lacey, *State Punishment*; Duff, *Punishment, Communication, and Community*; John Braithwaite & Philip Pettit, *Not Just Deserts: A Republican Theory of Criminal Justice* (Oxford: Clarendon Press, 1990), that an adequate justification of punishment must ultimately be part of a more fully developed political theory. On the other hand, of course, the integrated dualism put forth, if attractive, is to that extent an argument against any political theory with which it is incompatible.

⁶⁷ David Dolinko, “Retributivism, Consequentialism, and the Intrinsic Goodness of Punishment,” *Law and Philosophy* 16 (1997): 507-528, p. 520 n.37.

held by the wrongdoer; or, while conceding infringement, to maintain that the effectuation of a state of affairs in which wrongdoers receive the suffering they deserve justifies infringing the wrongdoers' rights, thus making it the case that the wrongdoers' rights are not violated. In fact, both types of reply can be discerned in retributivist writings. Moral forfeiture theories adopt the first approach explicitly. Theories that emphasize retributivism as a species of justice are perhaps best understood as adopting the second. After briefly assessing these two familiar retributivist lines of argument and finding them wanting, I will suggest a third alternative. This alternative, which we might call "right-rearticulation," pursues the ambition of moral forfeiture theories – to show that core cases of punishment need not infringe rights – but without the mysterious forfeiture mechanism.

2.

Forfeiture analysis starts by acknowledging that individuals do generally have a right that others not inflict suffering upon them (along, perhaps, with such other seemingly implicated rights as the right not to be deprived of liberty), which is to say that there exists such a thing as a right against the infliction of suffering. The forfeiture theorist then argues that wrongdoers forfeit that right by virtue of their wrongdoing.⁶⁸ By eschewing a need to demonstrate that there exists reason of overriding force sufficient to justify the infringement of a wrongdoer's right that his suffering, albeit deserved, not be inflicted upon him, the forfeiture route to J(P)(infliction)

⁶⁸ See, e.g., Alan H. Goldman, "The Paradox of Punishment," *Philosophy & Public Affairs* 9 (1979): 42-58; C.W. Morris, "Punishment and Loss of Moral Standing," *Canadian Journal of Philosophy* 21 (1991): 53-79.

proceeds, in essence, by cancellation.⁶⁹

Although superficially appealing, this claim runs into familiar and formidable difficulties, of which two stand out.⁷⁰ First, the mechanism by which forfeiture is effected is unclear. Surely A's violation of rights of B does not logically entail that A thereby forfeits any rights of his own. And any suggestion that A has chosen to relinquish any of his rights by offending against B must employ the notion of choice in a stipulated and tendentious sense. So it not wholly clear what can be said to explain why a wrongdoer forfeits rights beyond either asserting that it is so or conceiving rights as the product of a social contract that includes forfeiture among its terms. Second and yet more significantly, it appears exceedingly hard to specify just which rights are forfeited and for how long. While it is implausible to suppose that wrongdoers forfeit rights wholesale and for good, arguments for forfeitures more limited in time and scope often appear opportunistic and unsatisfactorily grounded.

3.

If the theory of moral forfeiture is unpromising, one might think the better alternative is to

⁶⁹ In contrast to the often unexamined assumption that the justification of punishment must proceed by override, see *supra* note [44], moral forfeiture theories have sometimes been thought attractive precisely for denying that punishment must be justified by override – if it is to be justified at all. See, e.g., Duff, *Punishment, Communication, and Community*, p. 15 (“[W]hile we can sometimes talk of rights as being justifiably overridden or infringed, this approach does not seem appropriate here: for someone whose rights are justifiably overridden or infringed is due an apology, or even compensation, whereas those who are justly punished surely have no such claim.”).

⁷⁰ See, e.g., Burgh, pp. 198-202; Warren Quinn, “The Right to Threaten and the Right to Punish,” *Philosophy & Public Affairs* 14 (1985): 327-373, pp. 331-334; Duff, *Punishment, Communication, and Community*, pp. 15-16.

concede that punishment infringes wrongdoers' rights while denying that any such infringement amounts to a violation. It is in this space that purported justifications of punishment that sound in "retributivist justice" might profitably be situated.

Notice that the retributivist claims endorsed thus far (in Section IV.A) take two forms that I have treated as equivalent: wrongdoers deserve to suffer, and the suffering of wrongdoers is an intrinsic good. In thus offering a retributivist response to $J_n(P_c)(\text{suffering})$, I have not needed to deploy the concept of "retributive justice." The familiar retributivist contention that retributivism is a species of *justice*, and not merely a label for a particular view of intrinsic goods – as in, "retributive justice demands that we inflict punishment in accord with an offender's just deserts" – might serve instead as a response to the demand for justification now under consideration, $J_n(P_c)(\text{infliction})$.

The state has an interest in bringing about various goods, where the strength of its interest is proportional to the importance of the good at issue. Those who doubt that the intrinsic good of a wrongdoer suffering (on account of and in proportion to his blameworthy wrongdoing) provides a reason of sufficient weight, or of the right sort, to render permissible the infringing of a right might think otherwise once justice is introduced as a mediating norm or value. Retributivists can now argue that a wrongdoer's suffering is demanded by justice, and therefore that what provides reason of sufficient weight to justify the infringement of a right is (at least in part) the state's obligation to see that justice is done, not (solely) the entailment or particularistic content of that obligation. Or, insofar as rights are conceived in terms of protected reasons,⁷¹ we

⁷¹ See, e.g., Joseph Raz, *Ethics in the Public Domain*, chs. 2 & 12.

might suppose that pursuit of the (mere) good that a wrongdoer suffer counts among the possibly countervailing reasons that the enjoyment of a right is specifically designed to exclude, whereas the possibility that infringing the right would further justice is a non-excluded reason. Happily, whether protected reasons do serve in the better conception of the function and structure of rights need not be resolved for our present limited purposes of understanding how the concept of retributive justice might be deployed as a response to $J_n(P_c)$ (infliction). Compare: Suppose that each of us has a right that the state not deprive us of property. Taxation infringes that right. But to the extent that a taxation scheme is designed to redistribute wealth in accord with the demands of distributive justice, our rights to property have not been violated. The pursuit of distributive justice justifies infringement of one's right against the deprivation of property. So too, perhaps, the pursuit of retributive justice justifies infringement of one's right against the intentional infliction of suffering.

This "retributive justice" argument for $J(P_c)$ (infliction), much like the consequentialist approach to which I adverted a few pages ago, implicitly proceeds by override: the pursuit of justice provides a reason to inflict suffering that outweighs (and is not excluded by) the rights-based reason not to.⁷² However it depends, of course, on the proposition that one's suffering on

⁷² One might suppose, to the contrary, that the proposition that justice demands the infliction of suffering upon A is incompatible with the proposition that A has a right against that infliction of suffering. On this view, the right in question might be better conceived as the right against the infliction of undeserved suffering, or something of this sort. And if this is so, then the argument sounding in retributive justice works by cancellation not override. The distributive justice comparison offers reason to doubt, however, that the incompatibility claim is sound. That distributive justice is served by redistributing property from A to B might justify the taking of A's property but does not entail that A lacks a right that his property not be taken. (To be sure, as Larry Sager has remarked to me, there is at least one circumstance in which the forceful

account of one's wrongdoing is a dictate of justice, not merely a good. But whether "retributive justice" is a genuine form of *justice* is vigorously contested. If the suffering a wrongdoer deserves is conceived comparatively, then retributive justice might be just a special case of distributive justice: the bad of suffering should be distributed among individuals in proportion to their relative degrees of wrongdoing. But the noncomparative conception of deserved suffering that I find much more faithful to retributive impulses (for example, it can support the conclusion that Cain deserves to suffer on account of his killing Abel, as the comparative conception seemingly cannot), is not easily assimilated to distributive justice, making the argument I have just sketched on behalf of retributivism dependent upon retributive justice being a genuine *and independent* form of justice. And that cannot be established by the sorts of argument employed in support of the more modest claim that the suffering of wrongdoers is an intrinsic good.⁷³

redistribution of property from A to B can be justified precisely on the ground that A has no right to the property taken – namely, when the Lockean proviso is not met. I assume here that distributive justice might demand more redistribution than the Lockean proviso could itself countenance.)

⁷³ There is a view about the relationship between desert and justice whereby it is analytically true that A's deserving X entails that, all else equal, justice demands that A be given X. See, e.g., Serena Olsaretti, "Introduction: Debating Desert and Justice," in *Desert and Justice* ed. Olsaretti (2003): p.8 (distinguishing between "telic" and "deontic" desert). I do not share this view, but need not argue against it here. I have thus far treated the desert claim (wrongdoers deserve to suffer) and the intrinsic good claim (it is an intrinsic good for wrongdoers to suffer) as equivalent. But I will deny this equivalence if challenged by those who believe that the desert claim entails the justice claim (justice demands that wrongdoers be made to suffer). My point here, then, would be to drive a wedge between the intrinsic good claim – which is all that the argument of Section IV.A needs – and the desert claim, not between the desert claim and the justice claim.

Still, whether my argument to this point be read as endorsing the intrinsic good claim and eschewing reliance on the desert and justice claims, or as endorsing the intrinsic good and desert

4.

In any event, a further difficulty confronts the retributivist-justice approach if that approach is conceived as a way to establish that wrongdoers' rights, albeit infringed, are not violated. There is no consensus regarding precisely what it means to infringe a right justifiably. On a common view, however, the non-violative infringement of a right still leaves a residue of some sort – paradigmatically, if is not exclusively, of compensation.⁷⁴ Yet it is far from obvious that fully justified cases of punishment do leave a residue; surely the state owes the wrongdoer neither compensation nor apology, for instance. Perhaps, then, moral forfeiture analysis, despite its own defects, was on the better track. With forfeiture theorists, that is, we might deny that wrongdoers have a right against the infliction of suffering and hence that P_c necessarily infringes any right. Put another way, we might wish to find other ways to make good on the cancellation approach to $J(P_c)(\text{infliction})$ that moral forfeiture theory embodies. The possibility I propose here takes the seemingly more extreme position that nobody has a right against the purposeful infliction of suffering, that there is no such right.

At first blush one might worry that such an approach would deny rights *tout court*. For if

claims and eschewing reliance on the justice claim, nothing in this section purports to come close to disproving the justice claim. In fact, I think retributivism *is* a plausible form of justice, independent of distributive justice. But the argument for that position can hardly be attempted in this space – partly because of its complexity and partly because components of that argument will turn out to be derivative of the arguments I advance in the immediately following section. At this point it is enough to observe that my retributivist response to $J_n(P_c)(\text{infliction})$ does not depend upon the concept of retributive justice but nor does it rely upon a rejection of the concept.

⁷⁴ This claim is familiarly illustrated by the hiker lost in the woods who breaks into your cabin for warmth and shelter. If he does not thereby violate your property rights, he does, precisely by virtue of having infringed your rights, owe you compensation.

any rights exist, the right against purposeful infliction of suffering would seem high on the list of plausible candidates. The worry gains force because, though the issue is undeniably complex, a categorical rejection of rights might seem to fit uncomfortably within a retributive framework. But to deny that there exists a right against the purposeful infliction of suffering is not to assert that there is no right in the vicinity. It's just to claim that this particular formulation gets the right wrong. The right in question – the right that the infliction of suffering implicates – might be more perspicuously rendered as, say, the right to be treated with respect as a person, a right that usually entails (inter alia) that one ought not to be intentionally made to suffer, but not always. The purported right not to be made to suffer states a useful rule of thumb, nothing more.

The route to this position starts from the now-familiar distinction between foundational and derivative rights, where the latter are thought to bear a constitutive or causal relation to the former: If A has a (foundational) right, against B, to X, A will also have a (derivative) right, against B, to Y, if Y is a component of X or a necessary means to X. On some accounts, the large panoply of rights recognized in conventional morality issue from a small number of foundational moral rights, perhaps one. As Hillel Steiner explains in a recent and helpful summary of the literature, theorists frequently purport to derive the diverse array of commonly recognized moral rights “from one unifying or underlying right entitling its holders to be secure in a certain broadly designated personal condition: well-being, autonomy, self-respect, and agency are among those most favored.”⁷⁵ Against this background, I offer two modifications.

⁷⁵ Hillel Steiner, “Moral Rights,” in *The Oxford Handbook of Ethical Theory* ed. David Copp (2006): 459-79, p. 474.

First and most importantly, the monistic foundational accounts are too partisan in attributing privileged status to one single aspect of personhood. Second and relatedly, we should recognize an additional category of derivative rights – rights whose relation to the foundational right is merely conventional and probabilistic.

Reduced to its core, the impulse that nourishes belief in the existence of moral rights is the idea that morality must attend to the distinctiveness of persons – that people’s interests not be aggregated in a fashion that fails to accord moral significance to their distinct personhood. As Steiner’s summary highlights, accounts that aim to reduce the diversity of rights to a single foundational right advance different and *competing* views of the particular aspect of personhood most deserving of respect. In contrast, the central thrust of my proposed foundational right to be treated with respect as a person (or, differently described, the foundational right to moral regard for one’s distinct personhood) is to deny that there is any general right answer to the question of which aspect of personhood must be uniquely respected, or of which aspect trumps the others when the normal way of respecting various aspects of personhood (well-being, autonomy, etc.) dictate conflicting courses of conduct. Instead, one’s right to respect as a person demands that others attend with seriousness to her distinct personhood, while recognizing that such serious attention and respect can cash out in any of a number of ways – a number of ways that is finite, but (often) greater than one.

A full defense of this purported right would, of course, require a paper of its own. Briefly, though, for this right to figure into a satisfactory retributivist argument, its proponents must, at the least: (1) establish that *if* this is the correct formulation of the right, then it plausibly

follows that core cases of punishment need not infringe any right of the offender;⁷⁶ and (2) provide reasons to believe that this *is* the correct formulation, mostly because it can account for other case-specific intuitions.

The idea that the infliction of deserved suffering does not, by itself, infringe an individual's right to respect as a person rests on the simple and familiar idea that such infliction respects human autonomy. This is not to endorse the far more robust claim, often advanced,⁷⁷ that *failing* to punish offenders is inconsistent with the respect and regard they are due, and therefore that the state is obligated to engage in retributive punishment. Rather, to repeat, the claim is that the right to respect as a person, and the correlative duty to treat persons with such respect, often fails to prescribe a uniquely acceptable course of conduct. People have preferences 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. We might say that the right to respect as a person does not obey the principle of distributivity across the aspects of personhood that have the capacity to guide respectful behavior. If A has a contractual right, against B, to

⁷⁶ I say "need not" rather than "do not" out of recognition that punishment can be (and often is) so brutal or degrading as to infringe – and violate – the right. Likewise, the right to respect as a person can be violated by the procedures employed prior to the infliction of punishment. But the justification sought here is theoretical. It is not a defense of punishment practices in any actual jurisdiction.

⁷⁷ See, e.g., Herbert Morris, "Persons and Punishment," *Monist* 52 (1968): 475-501; von Hirsch, *Censure and Sanctions*; Duff, *Punishment, Communication, and Community*, ch.2; Jeffrie G. Murphy, "Marxism and Retribution," *Philosophy & Public Affairs* 2 (1973): 217-43.

receive either an apple or an orange or a banana, it does not follow that any one of the following statements is true: A has a right against B to an apple; A has a right against B to an orange; A has a right against B to a banana. In just the same way, the proposed foundational right to respect as a person does not entail that a person has either a right that her preferences be respected or a right that her interests be respected or a right that her capacity for responsible choice be respected. What she has instead (to oversimplify) is a right that her preferences be respected or that her interests be respected or that her capacity for responsible choice be respected, etc. Given the ambiguity of the right's entailments, it cannot always prescribe a uniquely correct action when preferences, interests, deserts (and perhaps additional things of importance to personhood) pull apart or point in different directions.

Because wrongdoers experience suffering as a bad, it is plausible to assume 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.⁷⁸ 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. To again invoke the Razian idiom, conformity with the

⁷⁸ See Steven Sverdlik, "Motive and Rightness," *Ethics* 106 (1996) (arguing for the importance of motivations to act evaluations).

right is a function not only of guiding reasons but of explanatory reasons as well. To inflict suffering on the genuine belief that the suffering is deserved need not fail, then, to treat persons with appropriate respect.⁷⁹

Quite obviously, this is not to contend that any and all infliction of supposedly deserved suffering is rights-respecting. Most notably, forms of torture that seek to destroy the offender's dignity and capacity for autonomous action do seem incompatible with respect for personhood. The right-rearticulation approach therefore seems better able than the moral-forfeiture approach to support the common intuition that the infliction of torture does infringe an individual's rights.⁸⁰ Even more plainly rights-infringing would be infliction of suffering that is *not* genuinely

⁷⁹ One anonymous referee for this journal helpfully asks whether it follows from my account that the judge who improperly gives her friend a break by dismissing his traffic citation fails to accord the friend the right to respect due a person (in addition to whatever other wrongs such behavior instantiates). My answer is that it does not. If, for example, the judge believes that the friend would view the judge's failure to dismiss the citation as a betrayal, or simply could not afford the fine, then the judge's decision to dismiss the citation (albeit corrupt) would presumably be motivated by solicitude for meaningful aspects of the friend's personhood (aversion to pain, very broadly speaking) and is therefore not necessarily inconsistent with the duty of respect owed the friend. Here, though, is the key point of my analysis: even though the duty of respect might be satisfied by the judge's dismissal of her friend's citation, it does not *require* that conduct. If, for example, the judge believes that dismissing the citation would frustrate her friend's interest in living a life of moral integrity (even though the friend does not now realize it), she could satisfy her duty to accord the friend respect as a person by refusing to dismiss the citation. When, as is often the case, a range of action can be consistent with the right to respect as a person, then two possibly counter-intuitive conclusions follow: (1) A can respect B's right to respect as a person by Xing or by -Xing; and (2) A can, by Xing, either comply with or violate B's right to respect as a person. In both cases, all depends upon the character of A's motivation.

⁸⁰ Dolinko identifies only three possible responses that the retributivist can make to the infliction of torture as punishment (as opposed to as a means of eliciting information). She could bite the bullet (1) by conceding that state infliction of torture can be morally permissible. Or she could accept the common judgment that torture is morally impermissible, either (2) by denying

supposed to be deserved, which supports the widespread judgment that what I earlier termed degenerate cases of punishment are not justified.⁸¹

This argument, it bears emphasis, seeks to support J(P_c)(infliction) by reasoning fundamentally akin to the cancelling condition that supports J(P_c)(suffering). We ordinarily suspect that suffering is a bad, one we have moral reason not to bring about. But when that suffering is deserved, it is not a bad. We frequently say that people have a right against the intentional infliction of pain. But such a derivative right is just a common implication of a more abstract and foundational right – the right to be treated with respect as a person. When suffering is believed by a substantial certainty to be deserved,⁸² then the intentional infliction of proportional suffering accords the respect due personhood, and thus does not amount to a rights infringement that needs further justification. If this is so, then J(P_c)(infliction). Given the argument in Section IV.A for J(P_c)(suffering), we can conclude that J(P_c)(inflicts suffering).

that the torture of a person—even an individual who is himself guilty of torturing others—can ever be an intrinsic good, or (3) by contending that the wrongdoer’s right not to be tortured is not a right that can be forfeited. David Dolinko, “Retributivism, Consequentialism, and the Intrinsic Goodness of Punishment,” pp. 525-527. I am suggesting a fourth possibility: (4) it is an intrinsic good that a guilty person suffer in proportion to his blameworthiness, and bringing about such suffering does not infringe the guilty person’s rights insofar as the infliction of punishment treats him with the respect for personhood to which he has a right. Unlike response (1), this view concedes that torture (as punishment) is impermissible; unlike (2), it is not committed to denying that, in extreme cases, the state of affairs in which an individual suffers torture can be an intrinsic good; and unlike (3), it does not depend upon the selective forfeiture of rights.

⁸¹ See *supra* note [28].

⁸² We can see how this position would help support an argument for the “beyond a reasonable doubt” standard of proof, though developing and defending the argument cannot be attempted here. Very briefly, the idea is that a posture of proper respect demands that a decisionmaker or institution-builder appreciate the gravity of a false positive and thus create procedural protections to substantially reduce that possibility.

C.

The right-rearticulation approach I have sketched on behalf of retributivists is easily contrasted with the more familiar moral-forfeiture approach precisely in that it refrains from suggesting that the offender has forfeited any right. The offender has forfeited nothing. The offender, like everyone else, has a right to be treated with respect as a person. Under this right, however, the particular treatment that one is due is necessarily sensitive to his own conduct. The conditions that make punishment justifiable are thus built into the very content of the right; they are not grounds for snatching the right away.

That this articulation of the difference is indeed substantive, not merely semantic, becomes apparent when we shift attention from P_c to P_p and, in particular, from $J_n(P_c)$ (infliction) to $J_n(P_p)$ (infliction). It would seem obvious that moral forfeiture theories furnish no reply to the latter. An innocent person mistakenly believed to have offended cannot sensibly be taken to have forfeited any moral rights. So if the fact of an offender's forfeiture of a right is necessary to justify punishment of actual offenders against the fact that such punishment involves the intentional infliction of suffering, then nothing could justify punishment of non-offenders against the fact that such punishment involves the intentional infliction of suffering, and we should be forced to conclude that not- $J(P_p)$ (infliction), hence not- $J(P_p)$ (inflicts suffering).

But the analysis supplied above actually applies equally to P_p as to P_c . Respect is a quality of persons' attitudes (or attitudes-in-action) toward others. Whether B's conduct toward A honors A's right to respect necessarily depends, then, on relevant beliefs held by B about A. Now, which beliefs are relevant is a complex matter. They cannot include *all* those mistaken

beliefs that, if true, would render the conduct appropriately respectful. (At a minimum, they do not include the mistaken belief that, by virtue of some characteristic, such as being gay, disabled, Jewish, black, female, etc., A is not a holder of this very abstract right.) But they do include mistaken beliefs about brute facts concerning A's conduct, such as whether A did in fact shoot C. If the infliction of deserved suffering on an offender is consistent with his right to respect as a person (and does not infringe any other right he has), then the infliction of suffering on a putative offender that would be deserved if the putative were actual, likewise is not rights-infringing.⁸³ So $J(P_p)(\text{infliction})$ follows automatically from $J(P_c)(\text{infliction})$.⁸⁴

⁸³ As the discussion in text indicates, proper analysis of the question whether punishment infringes a right involves two two-value variables: whether the person punished (A) actually deserves the punishment inflicted (p), and whether the person inflicting the punishment (B) believes that the punishment is actually deserved. These independent variables combine to yield four possibilities: (1) A deserves p, and B believes that A deserves p; (2) A deserves p, and B believes that A does not deserve p; (3) A does not deserve p, and B believes that A deserves p; and (4) A does not deserve p, and B does not believe that A deserves p. The right-rearticulation and moral-forfeiture approaches agree both that, in the event that punishment is imposed, there is no rights-infringement in case (1) and that there is a rights-infringement in case (4). I have claimed that the approaches divide in case (3), where a right is infringed under the moral-forfeiture view but not under the right-rearticulation view. It seems to follow that the positions are reversed in case (2).

⁸⁴ Larry Alexander (in conversation) asks whether an innocent person who has been wrongly punished is owed anything if and when his innocence is discovered and, if so, whether this shows that the imposition of punishment does infringe rights in peripheral cases. Yes and no. I assume that the state is obligated to apologize to, and compensate, a person whom it has wrongly punished, and must, of course, terminate any punishment that is ongoing. But it does not follow that the state must have infringed a right, for positive duties are often triggered by an actor's causal responsibility for bad outcomes, even in the absence of a rights-infringement. Consider, for example, the duty to express remorse for injuring someone even non-negligently, and the duty to rescue that arises upon even a non-negligent creation of the peril. To be sure, there are differences between the cases of non-negligent causation of harm and of intentional punishment. But those differences can be overstated. After all, in the cases we have in mind, the state intentionally inflicts punishment but is non-negligent with respect to the critical fact that the

No similar move can be made for $J(P_p)$ (suffering). Even conceding that the suffering of a wrongdoer is not a bad (insofar as the suffering is on account of, and in proportion to, his blameworthy wrongdoing), the suffering of a person erroneously convicted of criminal wrongdoing is a bad, hence a reason against the infliction of suffering upon him. I have argued that whether conduct infringes a right, and thus whether the existence of the right constitutes a reason against the conduct, can be belief-dependent. The right to be treated with respect as a

person punished is innocent. Yet more importantly, the examples supplied do establish that we cannot blithely infer a rights violation from the duty to apologize or compensate.

Kent Greenawalt has pressed upon me the different worry that this account might not apply to cases in which the state knowingly imposes punishment in excess of a defendant's blameworthiness because, in order to protect against false exculpations or for other reasons of administrability, it refuses to extend the sort of defense that the defendant would need to rely upon in order to establish his absent or reduced blameworthiness. Such knowing punishment in excess of desert might be the consequence, for example, of statutory restrictions on the defenses of necessity, duress or insanity. More broadly, it might result from the fact that, in revealing contrast to what the Supreme Court has concluded is constitutionally mandated for capital sentencing, see *Eddings v. Oklahoma*, 455 U.S. 104 (1982), non-capital defendants are not entitled to put forth any and all evidence that might plausibly mitigate their blameworthiness.

It seems to me that cases of this sort sit at the border between what I have termed peripheral and degenerate cases of punishment. See *supra* note [28]. I am frankly uncertain how best to think about them. I am partly disposed to think of them as unjustifiable and thus most properly classified as degenerate: If the state can fairly be said to know (a complex question regarding how to attribute beliefs in institutional settings) that the statutory punishment would be in excess of a given defendant's blameworthiness but that the facts that make that so are not legally cognizable, it strikes me as plausible to insist that some state actor (prosecutor or judge) exercise her discretion to dismiss the case or to reduce the sentence, as appropriate. A system that effectively prevents such an exercise of discretion would not, on this view, be morally justifiable. If, in contrast, cases of the sort Greenawalt has in mind are morally justifiable – hence not properly classified as degenerate – then they would comprise a (likely small) subset of peripheral cases in which it is not the case, contrary to what I have concluded in text, that $J(P_p)$ (infliction) follows from $J(P_c)$ (infliction). In these cases, the defendant's right to respect as a person *is* infringed but possibly not violated.

person is an example. In contrast, whether a state of affairs is intrinsically bad or intrinsically good (or intrinsically neutral) is belief-independent.⁸⁵ Accordingly, $J_n(P_p)$ (suffering) cannot be met by cancellation. If $J(P_p)$ (suffering), the justification must proceed by override in the form of the good consequences that P_p produces.

Because P_p is not practically severable from P_c , in that we cannot eliminate P_p without eliminating P_c as well, the good consequences fairly attributable to P_p include all good consequences attributable to the institution of punishment as a whole. So the good consequences can include (what I have supposed to be) the good of bringing about a state of affairs in which deserved suffering obtains. But even if we grant to the retributivist the desert claim, it still would not appear to be sufficient for $J(P_p)$ (suffering) if we are to credit the common judgment that the undeserved suffering of the innocent is a bad of greater magnitude than the deserved suffering of the guilty is a good. And if so, then if $J(P_p)$ (suffering), that must be by virtue of good consequences beyond the causing of deserved suffering.⁸⁶ The frequent conclusion that

⁸⁵ More precisely (to accommodate non-realist metaethics), whether a given state of affairs is a *token* of the type of state that is intrinsically bad (or good or neutral) is belief-independent.

⁸⁶ This is not to assume that the benefits of a system of punishment are necessarily capable of outweighing the suffering of those punished in peripheral cases, let alone that the system of punishment operating at present in any Anglo-American jurisdiction is in fact consequentially justified against $J_n(P_p)$ (suffering). My point is twofold: first, that the good consequences the system produces must form part of any successful argument for $J(P_p)$ (suffering), and hence for $J(P)$ (inflicts suffering); and second, that those good consequences are not plausibly limited to the (supposed) good of producing the good of deserved suffering. (I am grateful to an anonymous referee for this journal for prodding me to make this point clear.)

retributivism cannot justify what I have termed peripheral cases of punishment⁸⁷ appears vindicated.

V. SUMMARY AND OBJECTIONS

According to the analysis put forth thus far, core cases of punishment are justified, in light of the suffering they inflict, by cancellation. When the person punished is morally responsible for the commission of an offense, then her suffering on account of that offense, and commensurate with her blameworthiness, simply is not a bad. Furthermore, the state's inflicting what it takes to be deserved suffering, for the sake of realizing such desert, does not infringe the offender's rights insofar as such infliction treats the offender with respect as a person. For these two reasons, neither the (mere) *causing* of suffering, nor its purposeful *infliction* constitutes a reason against the practice.

In peripheral cases of punishment – when, despite the state's genuine belief, the person punished is not morally responsible for the commission of an offense or not morally responsible to a degree commensurate with the suffering imposed – the fact that the putative offender experiences suffering does constitute a reason against the practice. In those cases, however, that reason can be overridden by the benefits that the institution of punishment produces – namely, (a) the deterrence of wrongdoing that is produced by a credible threat to punish the guilty, and (b) the variety of goods (including incapacitation and, arguably, the production of the intrinsic good

⁸⁷ See, e.g., Schedler, “Can Retributivists Support Legal Punishment” (arguing that retributivists' inability to justify peripheral cases of punishment renders them unable to justify punishment at all).

of deserved suffering) that are produced by the actual imposition of punishment against the guilty.

All this is to restate the justification for punishment thus far developed. This final part anticipates objections. It does so in the context of separately summarizing what I consider the account's three key features – that it is *tailored*, *dualist*, and *integrated*.

A.

First, the justification offered is tailored, or modest, in not purporting to justify criminal punishment all-things-considered. Instead, it responds narrowly to (both aspects of) what I have identified as the principal or conspicuous demand basis giving rise to the animating proposition that punishment is in need of justification – viz., that punishment constitutes the intentional infliction of suffering. It is not a justification of punishment, full stop; rather, it is a justification of punishment against this particular conjunction of reasons (or putative reasons) against it.

To throw the tailored conception of justification into sharper relief, consider a contrasting conception. Ironically (given that the conception of justification I am employing bears affinities, already mentioned, to his conception of desert), Joel Feinberg articulates a nearly polar opposite vision. In his view, “[o]ne justifies an institution . . . by demonstrating that it yields greater benefits and has fewer disadvantages than any of the proposed alternatives to it.”⁸⁸

A justification of punishment, accordingly,

would begin with an enumeration of all of the widespread and equally distributed benefits of the system of punishments. The case-maker would then ask his

⁸⁸ Joel Feinberg, “On Justifying Legal Punishment,” in *NOMOS III: Responsibility* ed. Carl J. Friedrich (1960): 152-67, p.158.

opponent to perform a series of experiments in his imagination, asking himself just what life would be like under each of the systems alternative to punishment, including no system at all. The case would then conclude with a comparison of the advantages and disadvantages of each alternative and demonstration that the balance of advantages is with the system of punishments.⁸⁹

We could choose to measure extant and future theories of punishment against this conception of justification. But we should expect retributivists to respond by denying that this captures the sort of justificatory enterprise they had taken themselves to be engaged in.⁹⁰ Moreover, this conception suffers from three deficiencies. First, it is unable to account for the signal role that the fact that punishment inflicts suffering has played in generating the proposition that punishment does in fact need justification. To the contrary, if, as Feinberg implies, punishment is in need of justification in the same way or for the same reason as are all institutions, then the justification demanded of punishment would seem to be (mere) general justification, which rings false. Second, it is insufficiently sensitive to the dialectical nature of practical reasoning, what J.R. Lucas called, in an under-appreciated essay, “the layout of arguments.”⁹¹ Third, Feinberg’s approach virtually ensures at the outset that we will be blind to

⁸⁹ Ibid., p. 161.

⁹⁰ Consider Michael Moore’s acknowledgment that to claim that “the moral desert of an offender is a *sufficient* reason to punish him or her” does not deny that “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.” Moore, *Placing Blame*, pp. 34-35. As his fuller discussion makes clear, Moore effectively treats *sufficiency* to be relational in the way that I claim for *justification*. See *supra* note __.

⁹¹ J.R. Lucas, “The Lay-out of Arguments,” in *Prescriptive Formality and Normative Rationality in Modern Legal Systems* eds. W. Krawietz et al. (1994): 285-95. Helpfully contrasting the structure of mathematical argument, Lucas observes that

whatever insights retributivism might prove capable of providing. For just as I have argued that even the tailored justification of peripheral cases of punishment must rely on consequentialist considerations, so much more obvious is retributivism's incapacity to provide a non-tailored justification of even core cases of punishment.⁹²

B.

The second feature of the account is that it claims necessary roles for both retributivism and consequentialism. Theories of this sort have attracted various labels – “hybrid,” “mixed,” “pluralist,” “compromise,” “intermediate,” and “synthetic” among them.⁹³ Although the first two are, I believe, the most common, they, like “compromise,” connote, at least to me, that the account to which they attach are neither one nor the other – neither truly retributivist, nor truly

[m]ost arguments, not only about morals, but in the humanities generally, are of a different, dialectical, form. A case is made out *prima facie* and objections to it are made. These objections may be rebutted, and counter-arguments countered. Sometimes the argument may be conclusive, so that there is no more to be said, and often some consideration is decisive so that we have little doubt in making up our minds. But often we do so leaving open the possibility that further considerations could emerge which would justify a change of mind, and often we reach a conclusion only tentatively Moral arguments do not proceed *more geometrico*. A geometrical argument is no argument at all unless it is absolutely watertight: a moral argument, together with most arguments in humanities, is full of possible holes, but nevertheless holds water unless someone succeeds in picking an actual hole. . . . Much therefore depends in moral argument on a possible hole not being an actual hole, on the other party's conceding the point rather than contesting it

Id. at 285, 290.

⁹² It is therefore no accident that, by setting out to capture “a *comprehensive* justification of punishment,” Husak, “Why Punish the Deserving?” pp. 447, 453 (emphasis added), Husak ends up finding retributivism unsatisfactory.

⁹³ Wood, p. 302 n.2.

consequentialist, but some new third possibility. In my view, the account presented here is both retributivist and consequentialist: Retributivism straightforwardly justifies part of punishment (the core); consequentialism straightforwardly justifies the other part (the periphery). To signify that this account has a both/and, rather than neither/nor, character, I prefer to call it “dualist.”⁹⁴

Though I might be criticized (wrongly, I think) for offering *only* a tailored justification of criminal punishment, nobody is likely to disagree with the characterization of this justification *as* tailored. The claim that my account views retributivist and consequentialist considerations as jointly necessary to provide the most satisfactory (tailored) justification for punishment stands on different footing. That this is even a fair description of the analysis might be questioned, rendering some elaboration appropriate.

1.

Here’s reason to worry that my reliance on retributivist desert might be superfluous as support for J(P)(inflicts suffering). My analysis has supposed that consequentialist considerations can, by override, support J(P_p)(suffering). A fortiori they can support J(P_c)(suffering) by override as well. Furthermore, these considerations might prove competent even to justify the infringement of individual rights, hence to support both J(P_c)(infliction) and J(P_p)(infliction). Admittedly, to establish this latter claim would require sustained argument, especially given vigorous disagreements among rights theorists regarding the types of considerations that can serve to justify an infringement thereby forestalling a violation. But even

⁹⁴ In favoring this nomenclature, I follow Wood, though he does not offer the reasons for his choice of terminology.

assuming that rights infringements cannot be justified by the goods they produce, almost all agree that, under appropriate circumstances, they can be justified by the need to prevent other rights violations. Possibly, then, punishment in core and peripheral cases can be justified against the fact that punishment infringes rights on the consequentialist ground that such infringements are necessary to prevent, via deterrence and incapacitation, future rights violations. Until this possibility is foreclosed, we are not entitled to conclude that retributivism is necessary for J(P)(infliction), in either core or peripheral cases. And because we have already supposed that J(P)(suffering) can be supported consequentially, it would follow that we cannot yet conclude that retributivism is necessary to support J(P)(inflicts suffering) in any respect.

This powerfully important objection throws into sharp relief the crux of the case for retributivism against consequentialism, which is, first, that consequentialism can only justify punishment by override whereas retributivism can claim to justify punishment by cancellation, and second, that justification by cancellation always enjoys conceptual or logical priority. So retributivists can concede without great sacrifice that the goods that punishment is reasonably thought to achieve can, by providing reasons of substantial weight in favor of the practice, justify punishment (in core and peripheral cases alike) against the fact that punishment causes suffering. And they can even concede, *arguendo*, that consequentialist override can justify core and peripheral cases against the fact (if true) that punishment infringes rights. In making these concessions, the retributivist is entitled to emphasize that her arguments proceed by cancellation. By contending that wrongdoers deserve suffering, retributivists deny that the fact of suffering constitutes a reason against punishment in core cases. And if the infliction of believed-deserved

suffering accords with an individual's right to respect as a person, retributivists can reasonably deny that punishment need infringe any right – in core and peripheral cases alike. The logical priority of the retributivist position should be clear: Once the putative reasons not to punish – that it causes the bad of suffering, and that it infringes rights – are shown not to be actual reasons, then we have no need to search for reasons to punish of greater weight; there is nothing to be overridden. So if the desert claim is sound, and if an effort to rearticulate the right claimed to be at issue can succeed, then it is fair to conclude that the best justification for punishment (tailored against the fact that it inflicts suffering) is indeed dualist.

2.

But *is* the desert claim sound? Admittedly, my account comes closer to assuming than to proving that wrongdoers deserve to suffer. My integration of retributivism and consequentialism is thus of a qualified sort: it is conditioned on the proposition that wrongdoers deserve to suffer.

Certainly it would be nicer not to have to assume the desert claim, but to have proved it. This, I have suggested, is more than can be hoped for. (Or, if I'm mistaken about that, it is more than can be hoped for from me.) So take this analysis as a philosophical explanation in the sense advocated by Nozick – less an argument that (integrated) dualism *is* true than an explanation of how it *could* be.⁹⁵ Even this, it seems to me, is something we should very much want to know. Retributivists and consequentialists have been locked in fierce debate for centuries, counting philosophers of the first rank on both sides. To insist on a univocal justification, retributivist or

⁹⁵ See Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press 1981), pp. 8-18.

consequentialist, is to dismiss entirely its traditional competitor. In light of the champions each view has won over the years, that is something we ought not to do cavalierly. This is not, of course, a matter of comity or etiquette. Rather, the fact that a large number of committed philosophers have embraced any particular view might reasonably be taken as some evidence (defeasible, to be sure) that there's something to it. We might therefore want to know the best case that can be made for a dualist account.

Nonetheless, to concede that my dualist justification of punishment assumes the desert claim might provoke a further objection – namely, that whether wrongdoers deserve to suffer (or, equivalently, whether the suffering of wrongdoers on account of their wrongdoing is an intrinsic good) is precisely the matter in dispute between retributivists and their critics. If everybody already agrees on what follows *if* wrongdoers deserve to suffer then my account is vacuous.

The premise, though, is false. Even assuming that wrongdoers deserve to suffer, the analysis in Section IV.C demonstrates that retributivism cannot by itself justify punishment, full stop, for it cannot support $J(P_p)(\text{suffering})$. Against retributivists, therefore, my account argues that consequentialism is a necessary component of $J(P)(\text{inflicts suffering})$. Moreover, as already explained, many contemporary anti-retributivists have argued strenuously that retributivism provides no part of a satisfactory justification for punishment, even assuming *arguendo* that wrongdoers deserve to suffer, because retributivists have been unable to explain why it would be permissible for the state to inflict suffering notwithstanding its being deserved. The retributivist argument by cancellation for $J(P_c)(\text{infliction})$ suggested in Section IV.B aims to rebut this objection. In sum, I have argued that (pure) retributivists have claimed that more follows from

the proposition that wrongdoers deserve to suffer than can be sustained and that (pure) consequentialists have granted to that proposition less than it earns.

C.

We can imagine a three-part distinction among retributivist, consequentialist, and dualist approaches sitting atop a taxonomy of theories of punishment. The dualist approaches themselves might then be usefully divided between the structured and the unstructured. Unstructured dualist accounts are purely additive or “mere conjunction” theories. Structured accounts, in contrast, can be further subdivided into the hierarchical (or “split level”) and the non-hierarchical. The dualist account presented here is structured yet non-hierarchical. Call it “integrated.” It offers something more analytically precise and satisfying than a claim that consequentialist and retributivist considerations each goes some of the distance in justifying criminal punishment but are jointly necessary to provide all the justificatory force required. At the same time, it demonstrates that a structured dualism need not be “split-level,” in which the two classes of considerations perform different types of justificatory functions.

Necessary to this integration was our initial division of punishment into core and peripheral cases. Yet you might say that this is an arbitrary way of dividing the terrain and that we ought instead to accept Hart’s division between the general justifying aim of the institution of punishment and principles of distribution of punishment. And, further, you might reasonably doubt that any part of my analysis undermines the common judgment that, once questions are

formulated in Hart's fashion, a consequentialist answer to the first seems nearly preordained.⁹⁶

For to inquire into the "general justifying *aim*" of a practice is no longer to ask for only its "moral justification," but to seek its "rational justification" as well (or, possibly, instead).⁹⁷ And while the effectuating of retributive justice could fairly count *among* the practice's rational justifications, it is unclear how, except by fiat, the retributivist can identify realization of this particular consequence as a justifying aim without allowing that the aim is, accordingly, to realize net good consequences generally.⁹⁸

I do not claim that my way of carving the complex phenomenon of punishment is preferable to Hart's. To the contrary, I am persuaded that no single way of analyzing punishment is natural or essential. The test of a proffered analysis is whether it helps illuminate that which might otherwise have lain obscured. The Hartian (and Rawlsian) distinction between justification for the social institution of punishment and justification for individual applications of punishment within that institution has the tendency to make retributivism assume, at best, a marginal or secondary role. When punishment itself is viewed as the activity in question rather than as the tail end of an institution whose leading edge consists of the threat of punishment, and when core and peripheral instances of the activity are distinguished, it seems easier to appreciate why retributivists believe both that the desert of those upon whom suffering is inflicted is

⁹⁶ Burgh, p. 197 n.7 ("It seems to be contemporary dogma that social institutions are to be justified in terms of their desirable consequences.").

⁹⁷ On the distinction, and how it applies to the justification of punishment, see Dolinko, "Some Thoughts about Retributivism," pp. 539-42.

⁹⁸ See generally Dolinko, "Retributivism, Consequentialism, and the Intrinsic Goodness of Punishment."

morally significant and that its significance cannot be adequately grasped by consequentialists who think in institutional terms. An analytical framework that separates core and peripheral cases of punishment is no truer than one that separates the general justifying aim of punishment from principles of distribution. But it is no less true either.⁹⁹

D.

Let us consider a final (yet logically primary) objection. Part IV argued that P_c can be justified against the fact that it inflicts suffering by the fuller fact that for an offender to suffer on account of, and commensurate with, his desert, is not a bad at all, and that the infliction of believed-deserved suffering is not, by itself, rights-infringing. The demand basis is thereby (twice) cancelled. This argument depends upon the argument, from Part II, that, customarily, to

⁹⁹ In conversation, Mark Greenberg has worried that separately justifying core and peripheral cases of punishment is a decidedly odd way to proceed. We do not, for example, separately justify core and peripheral cases of surgery; rather, we justify the practice or institution of surgery by reference to the goods that it produces (in the core cases), with the understanding that those goods must be sufficiently great to outweigh the bads that arise in the predictable (peripheral) cases when the operation goes awry.

Most practices are justified by the good consequences at which they aim. In such cases, if the distinction between core and peripheral cases of the practice is drawn (as I draw it in the case of punishment) between those cases the practice is aimed at, and those which arise by accident, we can anticipate in advance that the latter will be justified, if at all, on the grounds that the goods aimed at and secured in the core cases are sufficient to outweigh the bads produced as unavoidable byproduct. Such is what we anticipate in the surgery example, and also what we discover, making the strategy of dividing the practice into its core and peripheral cases entirely unilluminating. The reason why it turns out that the strategy of separating core from peripheral cases of *punishment* proves illuminating is precisely because and insofar as justification in the core cases proceeds by cancellation instead of by override. Insofar as the considerations that justify the core cases (against the demand bases on which I focus) are not themselves reasons to engage in the practice, they do not spill over, as it were, to justify the peripheral cases, making it necessary that other considerations (reasons) be brought to bear.

claim that any actual or contemplated conduct, X, is in need of justification means that “ordinarily” or “presumptively” one ought not to X, on account of some demand basis, y. It follows that some consideration, z, provides justification if, given z, it is not the case that, by virtue of y, one ought not to X. More generally, the argument put forth in Parts III and IV thus depends on a particular conception of what it means for something to require justification. I have argued that X can be justified either by supplying overriding reasons to X or by cancelling the supposed reason(s) not to X. But perhaps this is not so. Rather, it might be argued, justification can be supplied only by overriding reason, in which case an admission that punishment is in need of justification entails a rejection of the possibility that the demand basis can be cancelled. It might be objected, that is, that my entire argument rests on a failure accurately to capture what it means for conduct to be in need of justification.

On the competing view I have in mind to claim that any actual or contemplated conduct, X, is in need of justification means there exists reason against it.¹⁰⁰ If true that the only actions or activities needing justification are those opposed by reasons then to claim that X is in need of justification by virtue of y is simultaneously to deny that the reason, contained in the demand basis y, not to X is cancelled. Because a cancelled reason is not a reason, if the reason that purportedly makes X wrong were cancelled, then X could not need justification. So to endorse Jn(P) itself necessarily precludes that punishment could be rendered permissible by cancelling reasons not to punish. If reasons not to punish were cancelled, then Jn(P) would be false.

¹⁰⁰ I take John Gardner to be an advocate for this position. See *supra* note [33]. See also, e.g., John Gardner, “In Defence of Defences” p.1 (“What calls for justification, we can all presumably agree, is that to which there is some rational objection.”).

Perhaps this is the better way to think about justification. Perhaps clarity would be advanced were all to agree that only wrongs can stand in need of moral justification. Although skeptical,¹⁰¹ I am willing to be agnostic. What I deny is the empirical claim that this view of

¹⁰¹ Why should we think of demands for justification in this way, as necessarily resting on actual reasons as opposed to merely apparent ones? Very possibly, the answer lies in my earlier claims that an action is in need of justification only if “ordinarily or presumptively” one ought not to do it, and that justification is demanded if there exists “apparent or putative reason” against it. It is natural to translate these passages into a claim that an action needs moral justification if and only if it is *prima facie* wrong. And it is also natural to suspect that the “prima facie” that is here qualifying “wrong” is the same “prima facie” that Ross employed to qualify “duty.” If Ross’s notion of prima facie duty is as incoherent as some critics have maintained (see, e.g., John Searle, “*Prima Facie* Obligations,” in *Practical Reasoning* ed. Joseph Raz (Oxford: Oxford University Press, 1978): 81-90), then it might seem to follow that prima facie wrong is likewise untenable. Indeed, this appears to be Gardner’s view. Endorsing Searle’s criticism of Ross, Gardner has contended recently (now writing with Timothy Macklem) that “Prima-facie reasons are just reasons, prima-facie duties are just duties, and correspondingly, prima-facie wrongs are just wrongs.” Gardner & Macklem, “Reasons,” p. 467 n.34. If so, then the proposition (which I might be thought to endorse) that “what is in need of moral justification is a prima facie wrong” must become “what is in need of moral justification is a wrong.”

Allow me two points in response. First, Gardner’s reliance on Searle is misplaced, for Searle misunderstood Ross. Searle argued that the term was ambiguous and that neither of the two meanings Ross gave it was tenable. According to Searle, applying the qualifier “prima facie” to duty could mean either that the putative duty is not an actual duty, merely an apparent one, or that the duty, though actual, is overrideable rather than absolute. Given this ambiguity, and given difficulties with both conceptions, Searle proposed that we retire the term. But Ross intended “prima facie duty” to signal neither of the contrasts that Searle supposes (apparent as opposed to actual, or overrideable as opposed to absolute). No doubt Ross invited misinterpretation by using what he conceded was infelicitous language. Ross, *The Right and The Good*, p.20. Still, he is best read as aiming to mark, not a distinction between types of duty, but rather a distinction between stages of reasoning about duty. “[A]n obligation can change status or ‘mode’ while ‘all things are being considered,’ changing from a prima facie to an actual obligation during the consideration of evidence.” A. John Simmons, *Moral Principles and Political Obligations* (Princeton: Princeton University Press, 1979), p. 27. See also Philip Stratton-Lake, “Introduction,” in W.D. Ross, *The Right and the Good* (Oxford: Oxford University Press, 2002), pp. xxxiii-xxxviii; Lucas, “The Lay-out of Arguments,” pp. 285-87.

Second, whatever might fairly be said against the Rossian conception of prima facie duty,

what it means to furnish justification is embraced by all who take themselves to endorse $J_n(P)$. I deny, that is, that the following argument is valid: “Because you have agreed that punishment is in need of justification, you are foreclosed from arguing now that the reasons not to punish are cancelled.” Plainly, anyone who has acknowledged $J_n(P)$ but who now accepts the arguments, by cancellation, just sketched for $J(P_c)$ (suffering) and for $J(P)$ (infliction) may reply as follows:

In affirming $J_n(P)$, all I really intended to communicate was my agreement with the proposition that there exist apparent reasons not to punish, and that it is appropriately the burden of one who defends criminal punishment to explain why those reasons are not effective. If you prefer, let us say that punishment stands in need of pseudo-justification, and denominate this proposition $J'_n(P)$. If it is the case that only wrongs are in need of justification and therefore that justification can be supplied only by overriding reasons, not by cancelling conditions, then I now explicitly deny $J_n(P_c)$ and affirm in its stead $J'_n(P_c)$.

That this captures how retributivists often think about justification is suggested by the frequent criticism that retributivists argue as though punishment were not in need of justification at all.¹⁰² If only wrongs can be justified, then the critics may be right that retributivists do not think that punishment need be justified. But that does not mean, as the charge implies, that retributivists think nothing need be said in order to render punishment permissible. Surely they accept a burden of explaining the permissibility of punishment given the *apparent reasons* why it

it is critical to recall that my claim is only that X is in need of justification if and only if there exists actual or apparent reason against it. Plainly, we need not use the term “prima facie wrong” to convey the point. Still, it is useful to have a term to denote the stage in the dialectic of practical reasoning at which we are confronted with as-yet undefeated argument purporting to establish the wrongfulness of an action, which argument is potentially defeasible either by cancellation or by override, or by a combination of moves.

¹⁰² E.g., Hart, *Punishment and Responsibility*, p. 9; S.I. Benn, “An Approach to the Problems of Punishment,” *Philosophy* 33 (1958): 325-341.

is wrong. But they think those apparent reasons are not valid reasons, therefore need not be overridden. Critics are entitled to believe that, strictly speaking, such a position denies the need to provide justification. But this criticism rests on a narrow and contestable conception of what it means to demand justification. In any event, the seeming implication that retributivists see no need to provide *argument* is false.

CONCLUSION

A sympathetic or charitable understanding of contemporary retributivism requires one to recognize that its proponents are rarely purporting to provide an all-things-considered justification of punishment. More often, they are offering a tailored justification against the fact that punishment inflicts suffering and both of its aspects – that it causes the putative bad of suffering and that, by inflicting that suffering, it putatively infringes rights.

I have aimed to show that this is a successful justification on two plausible assumptions: first, that wrongdoers deserve to suffer on account of and in proportion to their blameworthy wrongdoing; and second, that any supposed general individual right against the infliction of suffering is better understood as only a usually sound implication – not a logical entailment – of some more abstract right such as the right to respect as a person. If so, then retributivism succeeds in justifying core cases of punishment (those in which the person punished is responsible for the charged offense and receives punishment proportionate to his blameworthiness) against the fact that it inflicts suffering. Consequentialist considerations might provide justification too, but the retributivist justification enjoys logical priority.

So far so good for retributivism. But retributivism lacks the resources to provide (in addition to an all-things-considered justification for the core cases) even a tailored justification for peripheral cases of punishment – those in which the imposition of punishment rests on a mistake regarding the defendant’s responsibility for the offense. So if peripheral cases are to be justified against the fact that punishment inflicts suffering, and if punishment simpliciter is to be justified all-things-considered (as, ultimately, we should want it to be), then consequentialist considerations must be relied upon. In this way, a fuller understanding of the logic of justification supports at least the framework for an integrated dualist justification of punishment.