

MITCHELL N. BERMAN

## REHABILITATING RETRIBUTIVISM

(Accepted 10 August 2012)

**ABSTRACT.** This review essay of Victor Tadros’s new book, “The Ends of Harm: The Moral Foundations of Criminal Law,” responds to Tadros’s energetic and sophisticated attacks on retributivist justifications for criminal punishment. I argue, in a nutshell, that those attacks fail. In defending retributivism, however, I also sketch original views on two questions that retributivism must address but that many or most retributivists have skated past. First, what do wrongdoers deserve – to suffer? to be punished? something else? Second, what does it mean for them to deserve it? That is, what is the normative force or significance of valid desert claims, either with respect to retributivist desert in particular or with respect to all forms of desert? Because the answers that this essay offers are preliminary, the essay also serves as a partial blueprint for further work by criminal law theorists with retributivist sympathies.

### I. INTRODUCTION

Victor Tadros’s elegant and erudite new book, *The Ends of Harm: The Moral Foundations of Criminal Law*,<sup>1</sup> is an original and important contribution to the philosophy of criminal law. I find much about Tadros’s novel anti-retributivist non-consequentialist justification for punishment congenial. Yet that’s easy for me. I’m a pluralist about punishment: I believe that punishment can be justified in principle on the strength of many disparate rationales. But Tadros? Not so much. In particular, he devotes considerable space and effort

---

\* Richard Dale Endowed Chair in Law, Professor of Philosophy, The University of Texas at Austin. For very helpful reactions to an earlier draft, I am grateful to audiences at the live book symposium sponsored by the Rutgers Institute of Law and Philosophy and at the Hebrew University Law and Philosophy Workshop. I also thank: David Enoch, Alex Guerrero, and Matt Lister for penetrating written comments; Kim Ferzan for inviting me to contribute to this symposium; Victor Tadros for stimulating conversations regarding his project and my response; and Danielle Wolfson for excellent research assistance.

<sup>1</sup> VICTOR TADROS, *THE ENDS OF HARM: THE MORAL FOUNDATIONS OF CRIMINAL LAW* (2011).

purporting to demonstrate that retributivism is both implausible and morally grotesque. I do not find these criticisms compelling. In this paper I seek to defend retributivism against Tadros's attacks.

I should rather say that I seek to defend a particular version of retributivism. Consequentialist justifications of punishment – what Tadros usefully re-labels 'instrumentalist' theories to signal that they need not embrace consequentialism as a comprehensive moral theory – represent a recognizable class of approaches. Individual versions vary from one another regarding the particular goods that punishment might secure and the mechanisms by which it secures those goods. The same is true of retributivism. Even if we limit consideration to those central or paradigmatic forms of retributivism that would justify punishment in terms of an offender's negative desert, particular accounts espouse different positions regarding, for example, just what it is that offenders deserve, in virtue of what they deserve it, and what justifies the state in endeavoring to realize those deserts.<sup>2</sup>

I believe that some of Tadros's criticisms are unpersuasive on their own terms, but that others do draw blood against many standard forms of retributivism. Accordingly, my effort to rehabilitate retributivism in the face of the many and often penetrating objections leveled in *The Ends of Harm* requires that I offer reinterpretations or refinements of some retributivist claims. In particular, I will propose new answers to two questions that are absolutely central to any retributivist justification of punishment: First, what is it that wrongdoers deserve? Second, what does it mean for them to deserve it? I do not have fully developed answers to either question, and what answers I do have cannot be fully elaborated and defended in this space. So I will just sketch the positions I find attractive and then show how they, or similar views in their vicinity, help address Tadros's challenges.

## II. TADROS ON RETRIBUTIVISM

Retributivism, according to Tadros, conjoins two core claims: '(a) that the suffering of offenders is good because deserved; and (b) that the state has a right to bring about deserved suffering'.<sup>3</sup> Although I

<sup>2</sup> See generally Mitchell N. Berman, *The Justification of Punishment*, in ANDREI MARMOR, ED., *THE ROUTLEDGE COMPANION TO PHILOSOPHY OF LAW* 141 (2012).

<sup>3</sup> Tadros, *supra* note 1, at 9.

will later refine this account just a bit, I think it plenty adequate to start. Against retributivism so conceived, Tadros advances a two-pronged attack.

The first prong aims to show that the thought experiments upon which retributivists frequently rely offer less support for their view than they assume. Here's a simple representative version. Imagine two worlds that contain some horrific wrongdoer – Hitler, let us suppose. In World One, he lives happily after committing his wrongs. In World Two he suffers pain or deprivation in response to, or on account or in consequence of, his offenses. In all other respects, the two worlds are identical. In particular, Hitler does not commit any different or worse acts in the two worlds, nor are any other would-be wrongdoers deterred from wrongdoing by the suffering that Hitler experiences in World Two or emboldened to do wrong by the happiness Hitler experiences in World One. Retributivists believe that World Two is better, or contains more value, than World One. Those who share that judgment, they say, are committed to the proposition that it is intrinsically good or valuable that wrongdoers suffer on account of their wrongdoing.

Tadros offers (in Chap. 3) two alternative explanations for the case-specific intuitions or judgments that thought experiments of this sort are designed to pump. First, Tadros allows that it is intrinsically good that wrongdoers recognize that what they have done is wrong. To be sure, in healthy individuals, this judgmental dimension of guilt almost invariably carries with it an affective dimension that consists of painful feelings of frustration and anxiety. But, contends Tadros, it is the judgment, not the bad feeling, that is intrinsically valuable, and that we should aim at. Second, insofar as some act of wrongdoing produces harms that are distributable, we should distribute them to the wrongdoer herself and not to innocents, because doing so respects the principle that unavoidable harms should fall, when possible, on those who have risked their occurrence by their own choices.

These are nice discussions, in my opinion, especially of guilt.<sup>4</sup> Some persons who lean toward retributivism on the basis of the familiar thought experiments might be persuaded by Tadros's

---

<sup>4</sup> I believe that the 'bad feelings' that Tadros attributes to 'the affective dimension of guilt' extend beyond guilt and are caused by more than the 'combination of anxiety and frustration' that Tadros identifies as the causal factors. But these quibbles need not detain us.

alternative explanations to reconsider. But Tadros fully acknowledges that his explorations of guilt and choice cannot demonstrate that retributivism is false. They only show that there are alternative explanations for the intuitions that retributivist thought experiments often produce. That is, they show, at most, that retributivism is not needed to explain the data, and that other explanations are available. While Tadros thinks his alternative explanations are better, not even he claims that they establish more than that.

To see their limited force, let us add detail to our two worlds. In World One, let us now suppose, our wrongdoer comes to believe that what he did was morally wrongful. But recalling his collegiate study of Nietzsche, he succeeds in resisting any accompanying bad feelings. His judgments of personal guilt lack an affective dimension. In World Two, he believes what he did was wrongful and he is wracked by guilt or remorse. Given the enormity of his wrongs, his anguish is great. Those who believe that World Two is still better than World One have reason (I do not say *conclusive* reason) to accept retributivist desert notwithstanding the alternative possibilities offered in Chapter 3.

Accordingly, in order to defeat retributivism, Tadros needs additional arguments. These – the second prong of his attack – he marshals in Chapter 4, ‘Against Desert’. Although the arguments can be individuated more finely, we can start with Tadros’s own four-part enumeration: First, moral desert depends upon conceptions of free will and responsibility that are not compatible with the truth of determinism. Second, if an individual’s negative desert could ground punishment practices of the type with which we are familiar, the ground or basis of such negative desert must be ‘something akin to criminal wrongdoing’,<sup>5</sup> but if well-being should track desert it is most plausible that it track the quality of a person’s character or actions over her life and not single discrete actions. Third, the retributivist claim that what wrongdoers deserve is suffering is either implausible or ‘barbaric’, depending upon the form that such suffering is thought to take. Fourth, even assuming that wrongdoers do deserve to suffer on account of discrete acts of wrongdoing, it is illegitimate for the state to coerce its citizens to supply the resources needed to bring about this particular good.

---

<sup>5</sup> Tadros, *supra* note 1, at 60.

In short, then, Tadros's four challenges, or sets of challenges, to retributivism are: that moral desert is incompatible with determinism and that determinism is true; that real-life institutions of criminal punishment cannot be reconciled with the 'whole life' view of a person's desert and that the 'whole life' view is true or most plausible; that suffering is never intrinsically good; and that it violates principles of political morality for the state to seek to inflict suffering even if deserved.

### III. RETRIBUTIVISM REFINED: TWO THESES ON DESERT

I will elaborate upon, and respond to, these criticisms in the next section. Here I sketch core aspects of the version of retributivism that I find most persuasive. That version emerges from, or is partially constituted by, the answers I would provide to two questions that all retributivists should address. Both questions concern desert.

#### A. *What Wrongdoers Deserve*

Since Feinberg, it has become orthodox to analyze desert as a triadic relationship in which some agent, A, deserves some treatment or state of affairs, O, on account of some quality or action, B.<sup>6</sup> Feinberg termed B the desert basis. I have called O the desert object.<sup>7</sup> Tadros recognizes that retributivists have different conceptions of what we might call 'the retributivist desert object', i.e., the treatment or state of affairs that wrongdoers deserve in virtue of their wrongdoing. Most commonly, he says, retributivists believe that what wrongdoers deserve is 'to suffer', by which he means, I think, something much like 'to experience pain or anguish'. I agree that that is the dominant retributivist view.<sup>8</sup> I do not, however, think it the best view. As I see it, and to a rough first approximation, what wrongdoers deserve on account of their wrongdoing is that their lives go less well. To a second pass (additional refinements are likely to be warranted), I propose the following 'relativist' conception of retributivist desert:

<sup>6</sup> Joel Feinberg, *Justice and Personal Desert*, in *DOING AND DESERVING: ESSAYS IN THE THEORY OF RESPONSIBILITY* (1970).

<sup>7</sup> See Mitchell N. Berman, *Two Kinds of Retributivism*, in R. A. DUFF & STUART GREEN EDs., *THE PHILOSOPHICAL FOUNDATIONS OF CRIMINAL LAW* 433, 437 (2011).

<sup>8</sup> See *id.* at 438.

RD-R:

An individual who engages in wrongdoing deserves that her life go less well than it otherwise would have gone, and less well in proportion to the blameworthiness of her wrongdoing, and that she understands that it goes less well as a consequence of her wrongdoing.

As already noted, a full elaboration and defense of RD-R must await a separate paper. At this stage I content myself with emphasizing two features about this conception of the retributivist desert object and offering a few thoughts in support of its plausibility.

The two features of importance are these. First, my proposed conception of the retributivist desert object – what it is that wrongdoers deserve – is catholic regarding the currency in which any given wrongdoer's life should go less well. It allows that the lives of some or many wrongdoers should go worse in the particular respect that they experience more suffering of varied sorts. But it also leaves room for the possibility that a wrongdoer's life should go less well in the respect that certain interests of the wrongdoer go unsatisfied or that certain liberties are denied her. More generally, it is compatible with any and all accounts of welfare, and potentially even with non-welfarist accounts of what it is for a life to go well or ill. As a pluralist about the constituents of human welfare, I am sympathetic to the view that a wrongdoer's life should go less well in different ways (and not solely to different degrees) depending upon the character or nature of the wrongdoing in which the individual has engaged. Naturally, much more needs to be said on this score to fill out the picture. In any event, I agree with Tadros that insofar as retributivists believe that wrongdoers generally, or some wrongdoer in particular, deserves *to suffer*, '[w]e need a reason to pick out *this* way in which we could make a person's life go worse amongst the range of ways in which we might do this'.<sup>9</sup>

Second and yet more significantly, my proposed desert object is inherently relative – relative, that is, to the individual's own counterfactual baseline. The suffering version of the retributivist desert object invites critics (and sympathizers) to wonder what the absolute level of suffering should be and how other features of the wrongdoer's life affect that level. On my account, a wrongdoer deserves

---

<sup>9</sup> Tadros, *supra* note 1, at 74.

that his life go less well than it would have gone had he not chosen to commit a wrong.<sup>10</sup> This dynamic and relative quality to the desert object I propose will have profound significance in meeting particular objections that Tadros levels.

With these two features of RD-R on the table, it remains to say something in its favor. Here, then, is ‘something’: If W is wrongful, then A should not W. All else equal, the state of affairs in which A Ws is less good than the state of affairs in which A does not W. If A Ws, then it would be good if his life goes less well as a result. It would be good if A’s wrongdoing produces bad consequences for him.

Although I think that is something, I admit that it is not much. My hope is that it rings plausible enough to warrant further investigation and amendment by retributivists, and to provoke anti-retributivists to offer counter-arguments. Schematically, I imagine counter-arguments falling within either of two classes, depending upon just how thoroughly the proponent objects to the proposition, more general than RD-R, that it is non-instrumentally better that the happiness or goodness of people’s lives is broadly responsive or sensitive to the moral character of their choices. Imagine two sets of worlds: (A) those in which people who behave as morality requires experience better lives than do people who act wrongfully, all else equal, and (B) those in which the quality of one’s lived experience is distributed *entirely randomly* with respect to the moral quality of people’s choices and actions. If (A) worlds were not at all better than (B) worlds just in virtue of this difference, then RD-R would almost surely be false. But to deny that there exists a good-making feature in (A) worlds not present in (B) worlds would, I think, be an unpalatable bullet for many opponents of RD-R to bite. The second class of

---

<sup>10</sup> Two important observations. First, because RD-R relies upon a counterfactual, it confronts the usual worries concerning the facticity of counterfactuals. Furthermore, RD-R as stated produces implausibly severe (mild) conclusions when the counterfactual life path would have led to unexpectedly bad (good) consequences for the wrongdoer. To accommodate both sets of worries, I rather suspect that the counterfactual baseline should be given an epistemic rather than ontic gloss. If so, a further question arises regarding the extent to which such an epistemic baseline should be subjectivized. These wrinkles require more attention.

Second, the construction ‘her life should go less well’ does not mean that the *entire remainder* of an individual’s life after committing wrongdoing W should be worse than the entire remainder would have been had she refrained from committing W. Her life should progress less well up to, but not beyond, the point at which the target diminution of her wellbeing, relative to her counterfactual baseline, is met. Therefore, RD-R is not necessarily frustrated when, for example, the anguish a wrongdoer experiences due to her wrongdoing prods her to change in ways that eventually produce for her a vastly better life than she would have experienced had she not committed the wrongdoing at all.

counter-arguments, accordingly, would aim to defeat RD-R without *categorically* denying that it is better for well-being to be sensitive to the moral quality of one's actions.

### B. *The Normative Force of Desert*

What, if anything, wrongdoers deserve is a question that must be addressed by retributivists. The second question to explore has wider significance, for it concerns desert generally, and not retributivist desert in particular. It is a question that should be answered by theorists who maintain, in any context, that A deserves O on account of B – whether O is suffering, happiness or a trophy, and whether B is wrongdoing, virtue, or having consumed the most hot dogs, with buns, in 10 minutes.<sup>11</sup> The question is: What does it *mean* to deserve O? Put differently, what relationships obtain between a valid claim about desert and other normative concepts? Because I am particularly interested in what consequences, if any, the fact that A deserves O has for the reasons for action, or duties, faced by some other agents, I will describe this as a question about the ‘normative force’ or ‘normative significance’ of a valid desert claim.

While Tadros properly recognizes how important it is for retributivists to specify a desert object, he does not expressly address this second question. That is a modest criticism, for this question escapes almost everybody who writes about retributivism, whether to praise or to bury. Now, Tadros appears to have an answer to this question, but because he does not seem to appreciate the importance of the question or the possibility of alternative answers, his answer is implicit. He appears to believe that a claim that some agent, A, deserves O on account of B is equivalent (or nearly so) to the claim that, given B, A's getting O would be intrinsically valuable. That is, Tadros appears to reduce a claim about desert to a claim about intrinsic value.

That too is a plausible and common view. (I have in past work assumed it myself).<sup>12</sup> But it is not the only view. The principal competing view in the desert literature maintains that desert is not only a

<sup>11</sup> The world record, set by Joey Chestnut at the 2009 Nathan's Famous Fourth of July International Hot Dog-Eating Contest, and tied by him in 2012, is 68.

<sup>12</sup> See Mitchell N. Berman, *Punishment and Justification*, 118 *ETHICS* 258, 269 (2008).



claim about value but a claim about justice.<sup>13</sup> In particular, it is often thought that A's deserving O on account of B entails that, given B, justice demands that A gets O. Of course, this view about desert's normative force must, to be complete, explicate what it means for justice to demand some outcome or treatment. I bracket that difficult question for now. It is enough for present purposes to note that justice and value likely bear differently on reasons for action. That is, the fact that some state of affairs is just (or that its negation is unjust) likely creates reasons of different scope and/or force than does the fact that that state of affairs is intrinsically valuable. Whether or not each of us always has pro tanto reason to promote the good (a controversial matter), it is more commonly held that we do always have reasons – and, very likely, obligations – to promote justice.

I have identified two candidates for the normative force of desert.<sup>14</sup> On the first, that A deserves O on account of B means no more than that, given B, A's getting O would be intrinsically good. On the second, it means that, given B, justice demands or requires that A gets O.<sup>15</sup> I suspect that both are mistaken: the first is too weak, the second is too strong. At any rate, the vast middle ground between these two accounts is worth exploring.<sup>16</sup>

<sup>13</sup> See, e.g., Samuel Scheffler, *Distributive Justice and Economic Desert*, in SERENA OLSARETTI ED., *DESERT AND JUSTICE* 69, 74 (2003) (presenting, as a common notion, the idea that 'justice consists, at least in part, in honoring claims of desert'). I find it striking how little explicit attention desert theorists have paid to the question of desert's normative force. It seems to me that many of the questions over which they obsess – like what sorts of things can be desert bases and whether desert is comparative – would be richly illuminated by a surer grasp of desert's relationship to value, reasons, and obligation. This too, of course, is a topic for another paper.

<sup>14</sup> A third common view associates desert with 'fittingness'. While I do not deny this association, I do not know what to make of it unless and until 'fittingness' is itself elucidated in terms of other familiar and better understood normative concepts. One commonly expressed worry is that, absent such elucidation, fittingness might be more an aesthetic notion than a normative one.

<sup>15</sup> The two possibilities I identify here are related, but non-identical, to Serena Olsaretti's distinction between 'telic' and 'deontic' desert. See Serena Olsaretti, *Introduction: Debating Desert and Justice*, in OLSARETTI, *supra* note 13, at 1, 8. Perhaps her distinction will prove to be the more useful; I am not certain. The critical nub is that 'desert sometimes is and sometimes is not thought of as related to obligations'. *Id.*

<sup>16</sup> Dissatisfaction with the two principal and competing unitary accounts of the normative force of desert might reasonably push us not to an alternative unitary account but, instead, to a pluralistic account under which some desert claims are claims only about value while others ground obligations. George Sher defends just such a view in his incisive and illuminating monograph. See GEORGE SHER, *DESERT* (1987). So a pluralistic account of desert's normative significance warrants serious consideration. For my money, though, Sher is more persuasive when arguing that the grounds or sources of valid desert claims are pluralistic than when arguing that the normative force is. In particular, he does not consider any intermediate possibilities like that reflected by NF-CJ. And given the desirability of finding a unitary account (it would help establish what provides coherence or commonality to desert claims that involve disparate bases, objects, and sources of value), I would not quickly abandon the search for it. In any event, even if Sher is right that no unitary account of normative force extends across all genuine desert claims, my narrow goals in this paper would be satisfied if something like NF-CJ were true of claims regarding retributive desert.

Consider, then, the following proposal in which desert bears *contingently* on duties of justice. I am far from claiming that this is the right account; I offer it to illustrate the possibility of accounts more nuanced and more attractive than those that conceptualize desert's normative force solely in terms of intrinsic value or solely in terms of (justice-driven) obligation:

NF-CJ:

That A deserves O on account of B means (a) given B, that A experience O (or that O obtain for A) is better than that A not experience O (or that O not obtain for A); and (b) if there is any agent or institution, X, with responsibility over the relevant domain, then X has a duty of justice to cause O to obtain for A.

Ultimately, of course, I will owe accounts of both (i) what determines whether an agent or institution has responsibility over a relevant domain and (ii) what 'duties of justice' mean. Unfortunately, I cannot provide nearly adequate answers to these very fair questions in this paper. Briefly, with respect to that first question, I am disposed to think that the domain over which an agent or institution may have responsibility sufficient to ground duties of justice could be defined either in terms of the desert basis or in terms of the desert object (or, perhaps, in terms of a conjunction of the two). For example, if it is true that A deserves to be rewarded for rescuing a cat from a tree, then the agent or institution, if any, that has duties of justice to effectuate a reward could be an agent or institution with responsibility over the welfare of cats generally or this cat in particular, or one with responsibility over the distribution of rewards. (Of course, to avoid circularity, whatever establishes that an agent does have responsibility over a domain must not be reducible to the fact that the agent has duties of justice to effectuate deserts within that domain.) With respect to the second question, it is enough to reiterate and rephrase what I said above: that Q is good has some normative force for us, but different normative force from the fact that justice demands that Q, or obligates us to promote Q.

Note that if some acceptable content can be given to the ideas that collectively constitute NF-CJ, then the second claim that Tadros attributes to retributivism ('that the state has a right to bring about deserved suffering') might not be as distinct from the first claim as he

and other anti-retributivists assume.<sup>17</sup> That is, to say, not merely that A's getting O would be good, but that A *deserves* O on account of B, might already go some distance to explaining why the state has a right (at a minimum!) to cause O to obtain for A. Put another way, the propriety of the state's acting to bring about a wrongdoer's negative desert might already be (partially) built into the fact of the wrongdoer's desert if desert means something more than intrinsically good, along the lines represented by NF-CJ.

#### IV. RETRIBUTIVISM AGAINST TADROS

In this final section, I explain why retributivism when amended or clarified to reflect RD-R and NF-CJ survives Tadros's challenges. Because I think the argument emerges more clearly this way, I will invert the order of Tadros's second and third lines of attack.

##### A. *Determinism*

For his first swipe at retributivism, Tadros outlines a variation on the familiar incompatibilist case against moral desert. For two reasons, I will skip past this argument. First, Tadros himself concedes that it is 'by no means decisive'.<sup>18</sup> Second, I believe that incompatibilist challenges to notions of blameworthiness and desert prove too much by throwing into grave doubt not only our backward-looking agent evaluations but also our forward-looking action-guiding moral norms. Conceivably, an incompatibilist embrace of determinism threatens all of morality, not just judgments about desert and the like. I'll grant that, in the event that this is not so, retributivism of the sort I defend here would depend upon the formulation of a satisfactory response to the challenge from determinism. This paper does not take up the burden of formulating such a response.

##### B. *The Desert Object*

'Suffering is normally bad', observes Tadros, sensibly. 'Retributivists claim that when suffering is deserved for wrongdoing it becomes

---

<sup>17</sup> For an important presentation of the anti-retributivist position that likewise assumes that the retributivist desert object is just an intrinsic good, see David Dolinko, *Some Thoughts About Retributivism*, 101 *ETHICS* 537 (1991).

<sup>18</sup> Tadros, *supra* note 1, at 60.

good, good in a way that can at least help to outweigh the badness of the wrongdoing'. Tadros aims to show that this claim 'is demonstrably false'.<sup>19</sup>

Tadros believes that there are two kinds of suffering, what we might term 'directed' or 'undirected'.<sup>20</sup> The paradigmatic example of undirected suffering is physical pain, but it can include as well things like undirected anxiety or depression. Directed suffering arises from the frustration of a desire or interest – for example, suffering at being unable to provide adequately for one's family or at failure to be awarded tenure. Punishments that take the form of deprivations of liberty generally result in directed suffering, and for a retributivist should *aim at* such suffering. Tadros's strategy for assessing whether directed suffering can ever be good is to assess seriatim whether directed suffering can be good when it arises from frustration of desires or interests to do, achieve, or experience good things (like having a good relationship with one's child), bad things (like raping), or neutral things. After claiming to have established that directed suffering can be good in none of these three cases, he then assesses whether undirected suffering can be good.

### 1. *Directed Suffering*

Despite some doubts, I will accept arguendo Tadros's contention that it is not good that anyone suffer at being prevented from doing good things, i.e., 'things of value'. As Tadros puts it, 'it is bad to make people feel bad about good things not being done'.<sup>21</sup> Even if that is so, I am unpersuaded by his arguments to the effect that it is not good that anyone suffer at being prevented from doing bad things. The contrary view, he explains, 'suggests that we can bring about the suffering of offenders by preventing them from doing bad things only by ensuring that they retain or develop their motivations to do those bad things'.<sup>22</sup> But I think that Tadros is here conflating the question of what *is* good with the distinct question of what *we should do* to ensure that that good state is effectuated. Surely a retributivist is entitled to agree that it is better that A entirely lack the motivation to rape than that he have it and suffer at being unable

<sup>19</sup> *Id.* at 73.

<sup>20</sup> I am drawing here on private email correspondence.

<sup>21</sup> Tadros, *supra* note 1, at 77.

<sup>22</sup> *Id.*

to rape. But that is not quite to the point. The retributivist claim presently under consideration is only that *if* A, a wrongdoer, does in fact have a motivation to rape then it is good that he suffer at being unable to satisfy that desire. To be sure, this clarification of this retributivist claim in no way demonstrates its correctness. Tadros himself finds it ‘odd’ to believe that, if an offender does have such motivations, then ‘the frustration that [he] experience[s] at not being able to act on them is good’.<sup>23</sup> Perhaps it is odd. Perhaps it is even mistaken. But if the claim is ‘demonstrably false’, it is not, as yet, demonstratedly false, which is what the reader might have understood Tadros to have promised.

Be that as it may, it is Tadros’s denial that it can be good for a wrongdoer to suffer at his being unable to do *neutral* things that I find most perplexing. His argument in support of that denial consists of just a single six-sentence paragraph, so warrants being quoted in full:

[1] Perhaps retributivists will claim that there are neutral things that the offender is prevented from doing, and that it is good that he suffers at failing to do things that are neither good nor bad. [2] But well-motivated people will not feel frustration at being unable to do things that are neither good nor bad. [3] It is difficult to believe that we should be pleased that they are motivated to do neutral things in order that they suffer through punishment. [4] Suppose that the offender is prevented from watching grass grow, and his watching grass grow is neither good nor bad. [5] Could we really think it good that he has the motivation to watch grass grow, and that he suffers at not being able to do so? [6] That doesn’t sound plausible.<sup>24</sup>

I do not think that this is a good argument. In my view, it is also not expressed with the same clarity that is generally characteristic of the book. So it will take a little unpacking to show why it fails. Here is the problem in a nutshell: Sentence [1] attributes to retributivists both a descriptive claim and a normative claim. The descriptive claim might possibly be interpreted in either of two ways. On either construal, however, it is unquestionably true. And the paragraph provides no persuasive argument against the normative claim. To be sure, the normative claim could be mistaken. Tadros surely believes that it is and he could be right. But if mistaken, that will be for reasons that Tadros does not unearth. Again, no demonstration of the falsity of the retributivist claim appears here.

<sup>23</sup> *Id.* at 77–78.

<sup>24</sup> *Id.* at 78 (bracketed numbers added).

That's the summary. I'll now elaborate, by walking through the paragraph more slowly.

Read literally, the descriptive claim in sentence [1] is vacuously true. Of course imprisonment prevents inmates from doing neutral things. The inmates of the New Jersey State Prison in Trenton are prevented by the facts of their incarceration from reciting limericks at the corner of High and Church in downtown Warwick, England on St. Casilda's Day. Yet if that activity isn't morally neutral it's hard to know what is. So I think we are safe to understand the descriptive claim that Tadros means to attribute or offer to retributivists as something like (a) imprisonment prevents an inmate from doing neutral things that he has a preference to do, or (b) imprisonment frustrates inmates (in part) by preventing them from doing neutral things. Interpretations (a) are (b) are equivalent insofar as frustration arises from being prevented from doing what one has a preference to do. The normative claim in sentence [1] would be that wrongdoers' suffering arising from the frustration at being prevented from doing neutral things is good.

At first blush, the descriptive claim that Tadros attributes to retributivists seems plainly true however it is construed: people do in fact have countless preferences or motivations for things that are neither good nor bad – watching tv, eating ice cream, playing angry birds, and so on – and they experience frustration-suffering when they are prevented from doing those things. Tadros does not squarely deny this. Instead, he claims (sentence [2]) that insofar as an individual does experience such frustration she is not well-motivated. This is doubly puzzling. First, the assertion seems plainly false unless non-standard meanings are given to the notions either of being 'well-motivated' or of 'things that are neither good nor bad'.<sup>25</sup> Second, even if true, it is not at all clear what follows. The burden for sentences [3]–[6] is to resolve these puzzles.

As a candidate for a neutral activity, let's focus on watching a fairly mindless but innocuous television program. Is this in fact a

---

<sup>25</sup> I invite the reader's introspection: does it seem to you (1) that you have preferences to do things that are neither good nor bad; (2) that you at least occasionally experience frustration-suffering when you are prevented from doing one or another of the neutral things that you would like to do; and (3) that neither (1) nor (2) undermines the adequacy of your motivation? Yep, me too.

neutral activity or is it a good activity?<sup>26</sup> Is it more like – to take Tadros’s proposed paradigms – watching grass grow or ‘hav[ing] a good relationship with [one’s] daughter’? Suppose the former. In that case, Tadros’s argument takes the form of a *reductio*: retributivists can believe (a) that the frustration-suffering a wrongdoer experiences by being prevented from watching tv is good only if they also believe (b) it is good that he is motivated to watch tv; yet (c) that latter belief would be absurd. I do not think that retributivists are committed to (b). First, retributivists need not believe that it is good that the wrongdoer be motivated by any neutral things. They might just take it as a fact about people – wrongdoers and others – that they are motivated by some neutral things. Second, even if Tadros is right that retributivists are committed to believing this is good, for otherwise the putative good of wrongdoers suffering from being prevented from doing the neutral things that they are motivated to do could not be realized, retributivists need not believe that it is good that any given wrongdoer is motivated to watch tv *in particular* (or this program in particular). They can believe that it is good that wrongdoers, like others, are motivated by, have preferences for, or gain satisfaction from *some* things that are neutral in themselves. In fact, it seems to me highly plausible that it is good that people have the capacity to gain satisfaction from things that are not themselves good. So I think Tadros is mistaken that, if ordinary activities like watching television are neutral, then retributivists are committed to anything absurd in maintaining that it is good when wrongdoers suffer by being prevented from pursuing these activities or interests.

Suppose now that watching tv is good. This is, I think, Tadros’s own view – not on that precise description, but on some more general description like ‘relaxing’ or ‘being amused’. On this assumption,<sup>27</sup> Tadros’s argument against retributivism amounts to

<sup>26</sup> If, under Tadros’s evaluative scheme, it is a bad activity, then the case serves only to reinforce the doubts I raised above regarding Tadros’s argument against the intrinsic goodness of a wrongdoer experiencing frustration-suffering by being prevented from doing bad things.

<sup>27</sup> I think that Tadros is wrong to think that, for purposes of his tripartite scheme, watching tv is like having a good relationship with one’s child. My having a good relationship with my children is impersonally good. It is good regardless of my preferences. My watching tv, in contrast, is not impersonally good. It is good for me insofar as it makes me happy and satisfies preferences that I happen to have and that are not themselves bad. The impersonal goodness in play here is only that agents are happy and have their (perhaps somewhat laundered) preferences satisfied. Put otherwise, in the tv case, unlike the child case, it is not that the thing that I wish to do is good, but that it is good that I get to do the thing that I wish to do. I do not press this point because, for the reasons I give in the text, I believe that Tadros fails to score a blow against retributivism whether watching tv is good or neutral.

the contention that any good of a wrongdoer experiencing frustration-suffering from being prevented from watching tv is always and necessarily outweighed by the bad produced by the fact that the prevention reduces the amount of relaxation or amusement in the world.<sup>28</sup> Retributivists are apt to believe otherwise. But I fail to see how Tadros has here advanced the ball against retributivists as opposed to having just restated what is in dispute.

## 2. *Undirected Suffering*

And what, lastly, about wrongdoers' experiencing undirected suffering, as by experiencing physical pain? For example, says Tadros, '[t]he pain caused by flogging, retributivists might think, is a good'. Tadros deems this 'a more plausible version of retributivism, in that pain involves no attitude towards either the good or the bad'.<sup>29</sup> And yet the entirety of his response to this 'more plausible version' consists of a single sentence: '[I]f this is the best version of retributivism, we must return to the familiar charge pressed against retributivists: that theirs is a barbaric view with a thin cloak of respect for the dignity of offenders'.

To play Michael Palin to Tadros's John Cleese, this is not an argument. It is contradiction adorned with an epithet. Furthermore, it is not clear that Tadros's charge of barbarism, even if true, hits the right target. Distinguish these two claims that a retributivist might possibly endorse: (a) wrongdoers deserve to suffer physical pain, and (b) our punishment practices should include the intentional infliction of such pain, as by administering flogging. The issue presently under consideration is whether (a) is true. But (a) cannot be disproven by establishing that (b) is false given that (a) does not entail (b). After all, as anti-retributivists like Tadros frequently (and rightly) emphasize, what wrongdoers deserve and what we should do all things considered are different questions. Yet the charge of barbarism is more

---

<sup>28</sup> Recall Tadros's statement, quoted earlier, that 'Retributivists claim that when suffering is deserved for wrongdoing it becomes good, good in a way that can at least help to outweigh the badness of the wrongdoing'. It is not clear what the second clause adds if, as I'd have thought, it is in the nature of goodness and badness as aggregative properties that they can 'help' to outweigh each other. If that is so, then the only controversial claim captured by that sentence appears before the comma. But the discussion in text suggests another possibility. Possibly, by 'badness of the wrongdoing', Tadros does not mean the badness of the wrongdoing that the wrongdoer committed, but rather the badness of the thing that must occur for a wrongdoer to experience suffering – here, the prevention of modest goods like relaxing or being amused.

<sup>29</sup> Tadros, *supra* note 1, at 78.



sensibly leveled at (b), not (a), on the plausible assumptions (i) that the intentional infliction of such pain is brutalizing or degrading to society, to those who administer such punishments, and/or to the offender himself, and (ii) that it is brutalization or degradation of this sort that qualifies a practice as barbaric.

Suppose, then, that (b) is barbaric (in something like the foregoing sense). Even if so, it does not follow either that retributivists are wrong to maintain that it is intrinsically good that wrongdoers experience pain-suffering or that they cannot justify state punishment. That the first possible conclusion does not follow is precisely what I established in the preceding paragraph. The second possible conclusion does not follow because it is open to retributivists to advocate a different mode of punishment as the best choice among available options all of which are suboptimal. Suppose that retributivists are correct that what wrongdoers deserve is to suffer physical pain and that Tadros is correct that we ought not to inflict that form of punishment on grounds of barbarism. If so, retributivists could fairly advocate that we choose a form of punishment that inflicts frustration-suffering (say, imprisonment), even though that is not precisely what the wrongdoer deserves, on the grounds that doing so is preferable, all things considered, both to allowing deserved suffering to go entirely unrealized and to acting barbarically by inflicting the particular suffering that is deserved. And as we have seen, Tadros does not establish either that wrongdoers do not suffer from having their preferences for neutral things frustrated or that such suffering is not good.<sup>30</sup>

### 3. *Beyond Suffering*

I have just provided two reasons why Tadros does not establish the falsity of the retributivist claim that the suffering of wrongdoers can be good. First, his argument to the effect that it cannot be good for wrongdoers to experience frustration-suffering at being prevented from doing things that they would wish to do and that are morally neutral in themselves (i.e., that are neither good nor bad from an impersonal point of view) does not succeed. (I also raised doubts about his argument as applied to things that are bad in themselves).

---

<sup>30</sup> And notice that I have said nothing about the prospects for a retributivism that espouses that wrongdoers deserve to experience undirected suffering of a sort that does not involve physical pain. I do not believe that Tadros addresses this possibility.

Second, his argument to the effect that it cannot be good for wrongdoers to experience pain-suffering is not an argument. It is time now to ask how Tadros's argument fares if we replace suffering as the retributivist desert object with RD-R.

Of course, one might simply deny that it is intrinsically good for wrongdoers' lives to go less well as a consequence of their blameworthy choices to engage in wrongdoing. The rejection of RD-R might be one upshot of a belief in determinism, for example, or it might follow from an embrace of welfarism. I have already acknowledged that I will put aside challenges that might arise from determinism. And, assuming determinism away, I find it hard to imagine a compelling basis for denying that the world would be better if people faced negative consequences for engaging in blameworthy wrongdoing than if their choices to engage in blameworthy wrongdoing had no impact on the quality of their lives – let alone if such choices tended to make their lives go better! Until a compelling argument is put on the table, I think we are entitled (though certainly not compelled) to provisionally suppose that RD-R or something much like it is true. What might nonetheless be said against retributivism?

One line of attack is to acknowledge that the goodness of a wrongdoer's life going ill as a consequence of his blameworthy wrongdoing provides a reason for others to seek to bring it about that the wrongdoer's life goes ill, but to maintain that this is not a permissible goal of a morally legitimate state. This is the subject of the next section. This section examines other possible arguments against retributivism assuming *arguendo* RD-R, arguments that belong more properly to ethics than to political theory. Happily, Tadros himself supplies or at least intimates what I think is likely the best such argument.

Here's the basic idea. Retributivism cannot rest content with advancing true claims about value; it is a set of theories designed to justify punishment. Yet even if RD-R is true, that would not provide any reason for anyone to do anything to try to bring about the diminution in a life's quality or value or goodness that RD-R deems good if the diminution to be desired was already secured or guaranteed absent any intentional human agency. In fact, many religious philosophies maintain that such diminution *is* secured without

human intervention. That's the great thing about the Afterlife and about karmic rebirth, for example.

Now, Tadros himself does not seek to drive a wedge between RD-R and retributivist justifications of human-administered punishment by invoking mysterious or supernatural *causal* mechanisms like God or karma. But he accomplishes the same end by suggesting an *analytic* or *conceptual* connection between wrongdoing and a life's going less well. As he puts it, if 'virtue is its own reward...., why is the same thing [mutatis mutandis] not true of vice?'<sup>31</sup> That is to say, '[t]he wrongdoer's life already [and necessarily] goes worse by being a source of disvalue'. If so, then Tadros can concede that it is intrinsically valuable that an individual's life go less well in virtue of, and in proportion to, the individual's blameworthy commission of wrongdoing – just as RD-R maintains – without conceding that this truth has any implications for responsive human action, including responsive human action by non-state actors. While it would have been good had Hitler's life gone unimaginably badly given his unimaginably horrible crimes, the argument goes, in fact it *did* go unimaginably badly just in virtue of his unimaginably horrible crimes.

There is much to be said in response. To start, that vice is its own punishment does not follow inescapably from the fact, if true, that virtue is its own reward. The structures of virtue and vice are asymmetric in various ways. For example, it is generally thought that whether a good act is virtuous depends upon the actor's actually being motivated by the good, whereas a bad act could be vicious even if the actor wasn't motivated by the bad. Nonetheless, I raise this possibility mostly to put it aside for I do believe that there is more than a grain of truth behind the idea that a vicious life is a bad one for the person who leads it. (As Tadros put it to me in conversation some years ago: could a parent imagine anything worse for her child than for that child to be a Hitler?) For several reasons, though, I think that that truth does not provide as much as the anti-retributivist needs in order to render RD-R inert as justification for retributivist punishment.

First, virtue and vice are characteristics of dispositions or patterns of behavior or the like, not of single actions. Even if living a vicious

---

<sup>31</sup> Tadros, *supra* note 1, at 70.

life is inherently terribly bad for the vicious, it does not follow that the analytic relationship Tadros invokes obtains between the quality of one's life and individual acts of wrongdoing.<sup>32</sup> Second and most importantly, even if a life necessarily goes badly (well) in proportion to the badness (goodness) of the actions that compose it, it does not follow that it goes fully as badly (well) as it would be good for it to do. In fact, that the inherent goodness of living a virtuous life is not all the goodness that it would be good for a virtuous person to enjoy is strongly suggested by the widespread belief that, even granting that 'virtue is its own reward', the virtuous *nonetheless* 'deserve to be happy'. It would be good, that is, if the goodness that characterizes a virtuous life were not just an intrinsic property of such a life but were to manifest experientially as well. But whether it does manifest experientially is contingent, not analytic. In just the same way, then, it remains highly plausible that the inherent badness that attaches to one's life just in virtue of one's engaging in wrongdoing is not all the badness that it would be good for the wrongdoing to produce for the wrongdoer.

### C. *The Desert Basis*

Suppose now that some acceptable retributivist desert object can be identified, like that the wrongdoer suffer or that his life go less well. Tadros maintains that retributivists cannot identify a plausible desert basis that would combine with the retributivist desert object to justify actual punishment practices that would strike us as familiar or remotely feasible. At least they cannot, says Tadros, consistent with W.D. Ross's view that 'the desert basis is virtue and vice over a whole life and... the deserved outcome is happiness and unhappiness over a whole life'.<sup>33</sup> Deeming Ross's 'whole life view' 'the most plausible view for the desert basis and the deserved outcome',

<sup>32</sup> Tadros invites this response by himself distinguishing wrongful choices and vice as desert bases. See *infra* Sect. IV.C.

<sup>33</sup> Tadros, *supra* note 1, at 68 (citing W.D. ROSS, *THE RIGHT AND THE GOOD* 56–64 (1930)). For another presentation of the 'whole life view' see Gertrude Ezorsky, *The Ethics of Punishment*, in GERTRUDE EZORSKY ED., *PHILOSOPHICAL PERSPECTIVES ON PUNISHMENT* xi, xxii–xxvii (1972).

Tadros contends that it raises at least three separate challenges for retributivists.<sup>34</sup> They must explain: (1) ‘why choice is so important to desert, and why it is only wrongful actions and not virtues and vices themselves’ that constitute the retributivist desert basis; (2) why any institution established to give wrongdoers the suffering they deserve should not also be committed to giving good-doers the happiness they deserve; and (3) why any suffering or punishment is deserved by wrongdoers who are already below their deserved ‘level of well-being’.<sup>35</sup>

I will leave a response to the second challenge for the next section. Here I respond to challenges (1) and (3). I explain why a sensible retributivist theory would focus on choice, and why a wrongdoer can have negative desert sufficient to warrant a punitive response regardless of how her pre-punishment level of well-being compares to what she would deserve on a whole life view.

In both cases, the explanations are straightforward given RD-R. That is to say, once we endorse a conception of the retributivist desert object pursuant to which what a wrongdoer deserves is that her life goes less well on account of her blameworthy wrongdoing than it would have gone otherwise, these two supposed challenges are no challenges at all. Why choice? Easy. What is thought good on RD-R is precisely that people’s wrongful choices don’t avail them and, to the contrary, make them worse off.

Why need retributivists not obsess over the extent to which a given wrongdoer sits above or below the welfare level that the aggregate of her life choices should warrant? Easy again. This challenge to retributivism gains all the force it possesses from the assumption that the retributivist desert object can be pegged to something external to the wrongdoer – presumably some absolute

---

<sup>34</sup> Before introducing the several challenges that he believes follow from the ‘whole life view’ of desert, Tadros raises a preliminary challenge. Observing (correctly) that retributivists disagree amongst themselves regarding whether the desert basis includes outcomes, he offers that ‘[w]e might doubt retributivism on the grounds that it lacks resources to generate compelling answers’ to the question of whether the fortuitous realization or non-realization of outcomes affect what is deserved. Tadros, *supra* note 1, at 67. Yes, we might, but we oughtn’t. While it is true that retributivism *qua* retributivism lacks resources to resolve this question, it is doubtful that retributivism, as a set of theories that would justify punishment, should be expected to possess such resources. More plausibly, retributivism, like any set of theories within applied moral theory, is entitled to draw from other philosophical domains. It follows that the failure of moral theory to have coalesced on views regarding the constituents of wrongdoing or blameworthiness is no more a mark against retributivism than is the fact that axiologists differ about value a mark against consequentialism as a comprehensive moral theory (or set of theories).

<sup>35</sup> Tadros, *supra* note 1, at 71.

norm of individual welfare, or a norm established by inter-personal comparisons. RD-R, however, pegs the desert object relative to an *intra*-personal standard. Whatever one's objective level of welfare may be at any moment in time (measured absolutely or by reference to standards determined by inter-personal comparisons), morality commands that she not engage in wrongdoing. Morality also provides – so says RD-R – that if she does engage in that wrongdoing, it would be better were her life to go less well as a consequence. Less well than *what*? Less well than *her* life would have otherwise gone.

Are the consequences of this view implausible? I do not see why. Consider two agents, A and B, contemplating the same wrongdoing, W, at t1. At t1, A possesses a higher level of welfare than does B. And to the extent that non-welfarist factors bear on the overall quality of one's life, A does better by this metric as well. It is also true (and not surprisingly) that if they both choose to refrain from W, A's welfare (defined expansively to encompass all respects in which a person's life can plausibly be understood to progress well or ill) will continue to be higher than B's.<sup>36</sup> According to a retributivist account that accepts RD-R, if both A and B choose to W, then it would be good if each enjoyed less welfare going forward than he or she would have otherwise. If each gets what her wrongdoing deserts her, A will end up still enjoying greater welfare than B. Retributivists should be willing without embarrassment to own that outcome. For one thing, it does not necessarily follow from RD-R that both A and B should experience an equal drop in welfare, measured absolutely. The magnitude of one's wrongdoing, and (especially) one's blameworthiness, for equivalent wrongdoing, can be sensitive to the actor's welfare level. More importantly, retributivism is a *localized* theory of justice, not a *comprehensive* theory. It is a theory about the grounds and contents of the good reasons we have to institute and operate a system of punishment for wrongdoing. It does not claim or aspire to be a theory about what we should do *taking all justice-relevant considerations into account*.

#### D. *Why the State?*

Suppose that my arguments to this point are correct. That still doesn't get retributivists all the way home, notes Tadros. 'For it is one thing to

<sup>36</sup> Recall my caution about counterfactual baselines. See *supra* note 10.

establish that it is good that wrongdoers suffer but it is another thing to establish that the state is permitted to bring about this good'.<sup>37</sup> The most powerful reason why the state might not be permitted to do so, he argues, is that doing so requires the state to expend significant resources that it extracts coercively from its citizens. But, says Tadros, citizens have the right 'not to have their resources used for the pursuit of a good that they might not wish to pursue'.<sup>38</sup>

I have two responses. First, the general limitation on legitimate state authority that Tadros invokes is doubtful. To be sure, Rawls and other liberal political theorists have argued that 'to justify a policy or institution under modern liberal conditions, we must be capable of showing that policy or institution to be compatible with a range of reasonable conceptions of the good'.<sup>39</sup> But that is a controversial position. For one thing, it would appear to delegitimize a range of state actions that are widely thought acceptable, including state funding of the arts and the provision of public parks. Understandably, then, Tadros eschews reliance on this strong principle, insisting that 'rejecting the idea that the state is permitted to punish on retributive grounds does not depend on accepting a neutralist version of liberalism of the kind that John Rawls and others defend'.<sup>40</sup>

But if the rejection of retributivist punishment does not depend on a Rawlsian neutralist version of liberalism, it is not pellucidly clear what principle it does depend on and what grounds that principle. Tadros's favored principle seems to be that it is illegitimate for the state 'to use significant resources of citizens' to pursue 'goods that are not grounded in the rights of others'.<sup>41</sup> However he provides no argument for this 'significance' condition and even expresses some doubt about it.<sup>42</sup> If the legitimacy of state coercion to pursue an impersonal good like the satisfaction of retributivist desert does not depend on the magnitude of the resources secured by means of such coercion, then it is no longer clear that Tadros avoids relying on neutralist liberal principles, as he claims.

<sup>37</sup> Tadros, *supra* note 1, at 79.

<sup>38</sup> *Id.*

<sup>39</sup> *Id.* at 80–81.

<sup>40</sup> *Id.* at 81.

<sup>41</sup> *Id.* at 81, 82.

<sup>42</sup> *Id.* at 83. ('Perhaps the state may spend modest resources simply on pursuing the good. I am not sure whether this is true. But the criminal justice system is very expensive'.)

Alternatively, if the legitimacy of state coercion to pursue retributivist desert does depend on just how expensive the criminal justice apparatus is, then I believe that retributivists can escape Tadros's charge by appealing to the difference between what I have elsewhere termed *tailored* and *all-things-considered* justification.<sup>43</sup> Plausibly, retributivism aims to justify punishment against the most universal, salient and (seemingly) forceful objection that it provokes – namely, that it is wrongful because it constitutes the intentional infliction of suffering or the intentional deprivation of important liberties. A retributivist justification can succeed on those terms even if it does not show that punishment is morally justified against *all* the objections that might be leveled against it, including its substantial cost in a modern incarcerative state. A theorist does not give up a retributivist *tailored* justification of punishment when supplementing it with instrumentalist accounts to furnish an *all-things-considered* justification.

My second response takes issue with Tadros's assumption that punishment, on a retributivist account, amounts to nothing more than an effort to bring about an impersonal good. That assumption depends, as we have seen, on reducing a claim about desert to a claim about intrinsic value. However, on the conceptual account of the normative force of desert that I have offered – NF-CJ – that any state of affairs or treatment is deserved means more than that it is good; it means too that an entity with domain-relevant responsibility has duties of justice to bring that desert object about. If that is right, then we must ask whether the state has responsibility over a domain in which RD-R falls.

Admittedly, many dots will have to be connected and vague concepts made more precise to fully defend an affirmative answer to this question. But the rudiments are plain. We might start by asking, not whether the state has a desert-relevant responsibility with regard to wrongdoing, but whether there exists *any agent* who has a desert-relevant responsibility with regard to wrongdoing. The answer is yes: the wrongdoer himself has such a responsibility. So if RD-R is true, a wrongdoer has a duty to ensure that, as a consequence of his wrongdoing, his life goes less well. At the same time, a wrongdoer's victim (and others with special relationships to the victim) has reason

---

<sup>43</sup> See Berman, *supra* note 12, at 262–266.



to ensure that the wrongdoer complies with the reasons of justice that he, the wrongdoer, has incurred by wronging the victim. And duties aside, surely the good of the wrongdoer's life going less well is one that the victim and her allies have reason to promote. If the state prohibits or limits such private enforcement by the victim (as by assuming a monopoly on the legitimate use of force), then the state thereby assumes responsibility to cause those desert objects to obtain. If something like this is correct, then the state does not pursue retributivist desert solely on the grounds that it is an impersonal good like art. It pursues retributivist desert because it has assumed responsibility over the domain of wrongdoing and what it deserts and because, in so doing, it has assumed obligations of justice to secure those deserts.

With this response in hand, we can address as well the objection I left open from last section: why any institution established to give wrongdoers the suffering they deserve should not also be committed to giving good-doers the happiness they deserve. Simple: That the state has assumed responsibility over responses to (certain types of) wrongdoing does not entail that it has also assumed responsibility over the domain of happiness or the domain of good deeds, or a domain defined by some conjunction of the two. In fact, there is no reason to believe that it has. Perhaps a deity has responsibility over these domains. Perhaps God has reasons of justice to bring about deserved happiness. But the state does not.

## V. CONCLUSION

The *Ends of Harm* is chock full of arresting insights and persuasive arguments in service of a novel instrumentalist justification for criminal punishment. In addition to developing and defending his intriguing affirmative account, however, Tadros avows an additional aim of 'unseat[ing] retributivism as the dominant view about punishment'.<sup>44</sup> I am skeptical that retributivism is nearly as dominant as Tadros, echoing many other recent observers, contends. But that is neither here nor there. The question is whether Tadros has given those who had been attracted to retributivism good new reasons for doubt. I think he has not. This is in part (but only in part) because

---

<sup>44</sup> Tadros, *supra* note 1, at 42.

retributivists should deny two claims that Tadros attributes to them. First, what wrongdoers deserve might not always be suffering, but instead (and more catholically and contextually) that their lives go less well. Second, what it means for wrongdoers to deserve this is not simply that such an outcome is good, but instead (and more robustly) that the state assumes a duty of justice to bring it about.

Punishment abolitionists think criminal punishment devilishly hard to justify. They are surely right – if what we hope to justify are the punishment institutions and practices actually operated by the legal regimes under which we all live. But the in-principle justification of some punishment for serious wrongs – say, imprisonment, under humane terms and conditions, of murderers, robbers, and rapists – rests on a multitude of independently adequate grounds. The moral justifiability of some punishment practices is overdetermined. While Tadros should be applauded for offering a rich and provocative new justificatory account, that novel account does not threaten retributivism.

*The University of Texas at Austin,  
727 E. Dean Keeton Street, Austin,  
TX 78705, USA  
E-mail: mberman@law.utexas.edu*