The Limits of Moral Argument: Reason and Conviction in Tadros' Philosophy of Punishment[1]

ERIC BLUMENSON
Suffolk University

ABSTRACT

For generations, philosophers of punishment have sought to revise or combine established theories of punishment in a way that could reconcile the utilitarian aims of punishment with the demands of deontological justice. Victor Tadros' recent work addresses the same problem, but answers it with an entirely original theory of punishment based on the duties criminals acquire by committing their crimes. The unexpected appearance of a new rationale for punishment has already inspired a robust dialogue between Tadros and his critics on many of the individual claims that, linked together, comprise his argument. This critique focuses instead on Tadros' theory as a whole and the methodology he uses to support it. It proposes that Tadros' argumentative strategy can't justify his rationale by virtue of (1) the extent and complexity of the moral reasoning he invokes, (2) the counter-intuitive results his theory produces in an array of specific cases, and (3) the superiority of a negative-retributivist account in which moral reasoning and intuitive judgments, and the principles and applications that flow from each, are coherent and mutually supportive. Victor Tadros responds to these arguments in an essay following this critique.

Keywords: Moral philosophy, punishment, criminal law theory, moral reasoning, Tadros

The philosophy of punishment covers enormous ground, but if one problem endures at its core, it is the conflict between the utilitarian aims of punishment and the demands of deontological justice. They seem mutually

1 I am particularly grateful to Victor Tadros for responding to this critique, to participants at the NYC Criminal Law Colloquium for their comments, and to Carol Steiker and Harvard Law School for arranging the symposium with Tadros at which I presented an early version of this paper. Thanks are also due to Jeff Alsdorf, Jeffery Atik, Claire Blumenson, Jeanie Fallon, Stanley Fisher, Gabriel Levin, Steven Nathanson, Derek Parfit, Pat Shin, Ken Simons, Rory Smead, Marion Smiley, Rayman Solomon, and Terrell Ussing for their comments and counsel.

exclusive in theory, and may often generate different sentences in practice. Finding a principled way to combine them has motivated generations of criminal law theorists. This challenge has served as a kind of conceptual Rosetta Stone that, if met, might clarify much of the field.

The philosopher Victor Tadros has applied his considerable skill to the problem and believes he has solved it. In his highly regarded 2011 book *The Ends of Harm* and in papers since refining some of his views,[2] Tadros offers a truly original justification for state punishment, and does so with impressive depth and clarity. The astonishing advent of a new moral account of punishment has already inspired three journals to publish symposia on its merits, and prompted punishment theorists to revisit widely varied areas of the field. Tadros and his critics have commenced a robust dialogue on many of the individual claims that, linked together, comprise his argument. I join that endeavor in Part 1(B), infra, enumerating what I see as weak or missing links in Tadros' argument.

Although I raise these substantive objections to particular claims, my principal interest is in the argument as a whole, and the methodology Tadros uses to construct it – subjects that are necessarily excluded from the piecemeal analyses that have occupied Tadros and his interlocutors to date. A central claim is that this methodology is out of balance: it places too much faith in conceptual argument and too little in intuitive moral judgment; its extreme reliance on distended chains of reasoning leaves no role for deeply held convictions about specific cases.[3] One can't reach an end-point of reflective equilibrium with such a methodology, and I believe this, more than any

2 Tadros has developed and in some respects revised his theory in response to critics (2012; 2013; 2015a) and in his Response to this critique (2015b).

3 Rawls distinguished three categories of normative beliefs that should play a role in the method of moral reasoning he called Wide Reflective Equilibrium: considered judgments about specific cases, moral principles and rules, and moral theories (Rawls 1971: 19-21, 48-51). The distinctions are orthogonal and overlapping, however, because we may also have considered judgments about moral principles and theories (Rawls 1999: 286, 289; also see Brun 2014). "Considered judgments" are akin but not identical to what others call "pre-theoretical convictions" or "moral intuitions." As an example, one might believe that slavery to be immoral (A) by virtue of an intuitive conviction that such is the case, or (B) by inference from other beliefs that yield that conclusion; judgment "A" would qualify as a considered judgment in Rawls' usage if it also satisfied certain epistemological safeguards – stability over time, relevant knowledge, impartiality, etc.. "Considered judgments about specific cases" thus supply two ingredients to moral inquiry, both integral to the method of reflective equilibrium: moral convictions regarding particular cases, which in their specificity can act as a check on more general principles and theories (and vice versa); and moral intuitions as provisionally credible sources of moral knowledge. Both dimensions of a moral belief are important to my critique. Regarding its degree of generality, section 3 argues that specific sentences the Duty View would generate are so unacceptable as to warrant rejecting the theory. Regarding its grounding, section 2 argues that intuitive moral judgments matter, and cast more doubt on Tadros' counter-intuitive conclusion than his distended chain of inferential reasoning can support.

particular weakness in Tadros' reasoning, accounts for the theory's failure to persuade the numerous commentators who have sought to pick his argument apart.

Methodological imbalance is not a problem for Tadros alone, but constitutes an occupational hazard for moral and legal philosophy generally, given that conceptual argument is at the heart of what philosophers and lawyers are trained to do. But Tadros' argument is an especially fruitful example with which to consider the use of reason, intuition and judgment in moral argumentation generally: it is precisely because Tadros' moral reasoning is so exhaustive, accomplished and transparent that its frailties and limits come into view.

This study proceeds as follows. Part 1 describes Tadros' justificatory theory of punishment, distilling his argument down to eleven sequential steps and identifying several weak or missing links among them. The balance of the article puts these piecemeal critiques aside and evaluates the methodology and strength of the argument as a whole. Part 2 argues that Tadros' argumentative strategy can't take him as far as he seeks to go, simply by virtue of the extent and complexity of the moral reasoning he invokes. Part 3 demonstrates the counter-intuitive results his theory would produce in an array of specific cases, and argues that results so at odds with strong and settled convictions count heavily against that theory. Finally, Part 4 demonstrates that an alternative – a form of negative retributivism -- remains more persuasive than Tadros' theory because it leaves us in a position of reflective equilibrium, in which moral reasoning and intuitive convictions, and the principles and applications that flow from each, are coherent and mutually supportive.

1. TADROS' THEORY OF PUNISHMENT

On one view, punishment is justified by the intrinsic goodness of a criminal's suffering in proportion to his desert. Tadros entirely rejects this idea; he believes that no one deserves to suffer and that suffering is never valuable in itself, whomever it afflicts. For Tadros, the only possible ground for punishing someone lies in its beneficial effects. His rationale for punishment is exclusively instrumental.

Tadros calls his philosophy of punishment "instrumentalist" rather than "consequentialist" because he wants to distinguish clearly between his justification of punishment and the comprehensive theory of morality known as consequentialism (2011: 25, 39-40). Were his theory consequentialist in the latter sense, he would face the familiar devastating objection: because

results are all that count, the theory could require imprisoning a mobster's innocent mother when there would be utility in doing so, and even her execution if it would deter more killings than the one it would inflict. This is unacceptable to Tadros, but so is the retributivist solution that grounds punishment in the offender's desert.

Tadros' third way is a hybrid position: an instrumentalist rationale for punishment situated within a non-consequentialist moral theory. He insists that the value of punishment lies in its deterrent impact, but also recognizes deontological side constraints on pursuing it – most importantly, the Means Principle prohibition on using a person merely as a means to another's benefit (13, 23). In Tadros' telling, this constraint places very stringent limits on government actions (so much so that only a libertarian state would seem to comply with it.[4]) We punish in order to reduce crime, but the Means Principle restricts its infliction to the guilty, Tadros claims, because only the guilty have a duty to submit to it. For that reason, Tadros calls his theory the "Duty View" of punishment.

Of course, everything depends on establishing that this duty exists, and that it derives from something other than desert. (Otherwise all Tadros has done is change words, substituting a "duty to suffer punishment" for "just deserts.") Tadros' starting point is the example of one person attacking another: uncontroversially, the assailant is morally liable to be harmed by defensive force. Tadros then argues that if the assailant completes the crime, his liability to suffer harm persists, extending to a series of residual and remedial duties that culminate in a duty to deter crime by submitting to punishment.

4 Tadros argues that the state may not tax people to finance retributive punishments, both because (1) liberal neutrality would rule out compelling people to finance a controversial, non-neutral conception of the good, and in any case, (2) citizens are not bound to expend resources to pursue the good unless they have a duty to do so (2011: 79-83). He argues that measures protecting people from crimes are not subject to either objection, because citizens have rights to security that create co-relative duties to provide it (82-3). By contrast, he classifies retributive justice as an impersonal good and "it is much more difficult to justify forcing a person to make a contribution to the pursuit of goods that are not grounded in the rights of others," like the promotion of natural beauty, because each citizen is entitled not to pursue them (81).

Of course, this objection would eliminate large areas of government funding well beyond the promotion of natural beauty or the arts. Massive infrastructure projects like America's rural electrification project in the '30s or interstate highway system of the '50s are also not "grounded in the rights of others." Even funding such things as health care or occupational safety would seem to require a showing that people have rights to them and co-relative duties to provide them through the state. For that reason, the implications of Tadros' argument seem to approach the strict libertarianism Robert Nozick (1974) propounded, and bring to mind Nozick's famous claim that redistributive taxation is akin to forced labor. Compare Nozick 1974: 169 with Tadros 2011: 79, wherein Tadros argues that to "use resources that a person produces for the pursuit of [a] goal is perhaps not as coercive as forcing them to work for the sake of that goal, but the difference is not terribly significant."

That's a very long road to travel. How does Tadros get all the way from the permissibility of defensive force to the permissibility of state punishment? A highly distilled and simplified version of Tadros' argument as I understand it consists of the following multiple-step progression.

1.1 Tadros' Argument in Eleven Steps

Liability to be harmed by defensive force

I. All persons have a moral duty to refrain from wrongful aggression against others.

II. If wrongful aggressor A commences an attack upon victim V, A has a residual duty to prevent its completion or harmful impact on V if possible. A's obligation includes incurring a proportionate degree of harm if necessary to thwart the crime.

Note: A's obligation to incur harm does not arise because he deserves it. It stems in part from a principle of distributive justice, the Choice Principle. On this principle, if someone must suffer, it is better that it be a person whose choice created the situation than someone merely trapped in it. (Whether it is "better" prudentially because we all have reason to value choice, or morally because it is fairer, neither view treats deserved suffering as good.) Here it was A's choice to attack V that made the threat of harm inevitable (2011: 56).[5]

III. V (or a third person) may enforce A's duty to avert her threat by using defensive force against A that inflicts no more harm than A would have been liable to suffer in discharging her own duty to avert her threat.

Residual duty to compensate victim by protecting against other crimes

IV. If A's attack succeeds and harms V, A has a residual duty to provide a remedy to the victim.

V. The remedy A must provide V is protection from future crime, even at significant cost to herself, as long as that harm is (a) no more than A was liable to suffer from V defending himself at the time of the crime, and (b) proportional to the harm it would prevent.

5 Tadros (2011: 56) says that while the opportunity to avoid being harmed will often coincide with culpability, it is the element of choice rather than desert that is basic.

The obligation to submit to defensive force appears over-determined in Tadros' theory. Whether the Choice Principle is necessary to his argument is left unclear given his sporadic reliance on the aggressor's breach of her duty of non-aggression to justify, by itself, the residual duty to suffer defensive force that follows. If it is the latter, there is a question whether forcing the aggressor to do what he had a duty to do can fulfill the duty, because it may completely bypass the aggressor's agency. (See Duff 2013: 116-117.)

Note: suppose A_1 has just shot V when A_2 arrives on the scene and independently attempts to shoot V. V may use A_1 as a protective shield against A_2's attack even though A_1 was not responsible for it, based on A_1's remedial protective duty. While this uses the aggressor as a means, it does not violate the Means Principle because she has a duty-based liability to suffer that harm.

Regarding the specification of protection as the remedy, Tadros argues that monetary compensation is ordinarily an inadequate remedy for a serious crime, and often unavailable in any event, so something more akin to specific performance is required of the aggressor: to protect the victim against a future crime of similar gravity (2011: 2, 277-78).

VI. On efficiency grounds, criminal wrongdoers are obligated to pool their protective duties and take responsibility for protecting each other's victims (193-94, 280).

Note: consider two assailants, each responsible to protect her particular victim from a future crime. If each assailant is unable to protect the person she attacked, but is able to protect the other's victim, both have an enforceable duty to do so. In the previous example, A_1 was liable to be used as a shield to protect V from A_2. As a result of this implied exchange, A_1 may now be used as a shield to protect someone else.

Duty to submit to punishment

VII. If punishing wrongful aggressors can deter crimes against the victims they are responsible to protect, they have a duty to submit to punishment, subject to the same limits as step #5 (279-80, 291).

State's exclusive right and responsibility to punish

VIII. The state may enforce a wrongful aggressor's duty to submit to punishment by punishing her (395).

IX. Because all citizens have a modest duty of mutual protection, all victims are obligated to use their right to protection-through-punishment to protect others as well as themselves (298).

X. For reasons of prudence, effectiveness and fairness, the state is best able to fulfill the victim's duty to punish. Therefore, all victims have an obligation to transfer their rights to punitive compensation to the state (297-99, 304-05).

XI. As the exclusive instrument of enforcement of both citizens' and aggressors' protective duties, the state is obligated to punish wrongful aggressors (293, 299-305).

Hence the core principle of the Duty View: The state has the exclusive right and responsibility to punish a wrongful aggressor for the purpose of general deterrence, provided that (a) the harm the punishment inflicts is proportional to its beneficial consequences, and (b) does not exceed that which the aggressor was liable to suffer in order to avert his crime.

1.2 Some Weak Links in the Argument

Critics who quarrel with a particular step in an argumentative chain often assume that the argument is only "as strong as its weakest link." In fact, as the next section argues, an inductive chain of reasoning is not as strong as any of its individual links, however weak or strong, and this problem may prove fatal to an argument as distended as Tadros'. But first, with Tadros' individual claims now in mind, it is worth noting some particularly weak or missing links in that argumentative chain.

Step 5 - The prior step has established that if a wrongdoer does not thwart the crime he started, he must do the next best thing: he must provide a remedy to the victim. Step 5 specifies that "[t]he appropriate way to remedy that wrong is by providing protection to victims and other citizens against future harm." (2011: 2). But that claim dismisses alternative types of rectification – monetary payment in a theft case, for example, or surgery and long term care in a maiming case. This move has spurred a fair amount of argument between Tadros and his critics which I need not repeat here, other than to note that nothing in Tadros' further arguments obviates the two problems noted in the margin.[6]

Step 6 - This step seeks to establish the aggressor's obligations to people he has never threatened: he has a qualified duty to exchange responsibilities with other wrongdoers, so that each victim will be protected by somebody. The duty to pool responsibilities is contingent on the greater effectiveness

6 First, mandating a protective duty rather than permitting monetary or other compensation deprives the victim of any choice in assessing how she might best overcome the damages she suffered. In a theory largely concerned with assuring respect for the moral status of autonomous persons, why should the state arrogate this choice to itself? (See Tanguay-Renaud 2013: 154; Ferzan 2013: 193-94). Responding to this criticism, Tadros has argued that leaving this choice with the victim "may lead her to violate the duty of protection that she owes to others." (2013: 307). As I argue later in this section, it is more likely that the state will violate its duty to protect the victim's compensatory right by aggregating it with all others and trading off among them on utilitarian grounds.

Second, the residual duty is supposed to be the next-best thing to fulfillment of the original duty, and it is far from clear that the victim would be better off with protection from a possible future crime than with a remedy aimed at mitigating or compensating for damages the victim has already suffered (Tanguay-Renaud 2013; Ferzan 2013). Tadros assumes monetary compensation is inadequate, but a remedy that is contingent on the possibility of future victimization may be more so.

of doing so. Suppose A assaulted V, and now wants to fulfill her protective obligation by purchasing a guard house on V's street and spending hours a day there insuring V's safety. Assuming A's work is sufficiently onerous so that it not only provides the protection but also exacts the degree of harm A could be required to bear, is A then exempt from the exchange, and thereby lacking the special duties to protect other victims that would ground A's liability to punishment? If so, punishment will be imposed unequally, based on arbitrary factors like the wrongdoer's wealth or the victim's age.

Step 7 - Step 7 specifies that the wrongdoer's protective duty is to submit to punishment. Having already established the wrongdoer's duty to avert someone else's threat – serving as a shield, for example - Tadros says, "If punishing each wrongdoer can protect other victims of crime from future offending, each wrongdoer ought to accept that he must be punished. (Tadros 2011: 280) Assume punishment *can* protect victims from future crime via deterrence; note that Step 7 still doesn't establish that wrongdoers are restricted to fulfilling their protective duty by that means when there are other effective methods available. Punishment is only one of many methods of deterrence, and deterrence is only one of many methods of protecting people from crime. We may achieve deterrence without punishment by increasing the neighborhood police presence, and we may reduce crime without deterrence through social work, job training, or other programs that ameliorate criminogenic conditions like poverty; so obligating offenders to undergo or help finance such programs might also fulfill their protective obligations. Given the suffering that punishment inflicts and Tadros condemns, he needs a persuasive argument to bypass such non-incarcerative alternatives.

As this step makes clear, Tadros' justification for punishment is entirely contingent on the effectiveness of deterrent punishments, a relationship that is notoriously contested and hard to determine; and if the required deterrent value is present, contingent as well on what other consequences might accrue, as Tadros recognizes (2011: 40; also see Ch. 15, sec. V discussing what consequences may properly count in determining proportionality; and 30, 338, 348, 352-53).[7]

Steps 9 and 10 - Here Tadros seeks to transform the victim's individual right to protection-through-punishment into a collective right exercised exclusively by the state. In step 9, he establishes that victims have a duty to use their right to punish so as to protect others as well as themselves, based on the modest duty all citizens have to provide mutual assistance to each

7 America's mass incarceration policy arguably resulted from an unduly narrow focus on crime control, to the exclusion of the social damage that would result from the removal of vast numbers of men from their communities, the diversion of resources away from policies that might ameliorate criminogenic conditions, etc..

other ("easy rescue"). Step 10 then argues that the victim's duty to punish includes the obligation to authorize others to punish if they will do so more effectively and fairly, and that because the state is such an agent, the victim must transfer her individual compensatory right to protection-through-punishment to the state to enforce (301; also see 297-99, 302-07).

One difficulty with this two-step argument is that it depends on a duty of mutual assistance that can't support it. At the least, the duty of mutual assistance morally (though not legally) requires a passerby to undertake relatively costless rescues, like calling 911 or throwing a lifejacket to a flailing swimmer. Does it also require that a victim's compensatory rights be transferred to the state and transformed into a system of deterrent punishment benefitting victims and non-victims alike, as Tadros argues? That strikes me as a bridge too far.

Even if the duty of mutual assistance could justify some loss in benefits, it cannot justify the loss of the victim's right to the remainder. Although Step 10 is framed as a matter of more effective enforcement of the rights and duties of each victim, the state can only enforce them in the aggregate. My right to compensation for theft, and your right to compensation for torture, and all other victims' compensatory rights, become subject to trade offs based on factors such as which punishments of which kinds of crimes will have the "biggest bang for the buck." Rights may be defeasible, but they cannot not be subject to such maximizing cost-benefit calculations and remain rights.

Indeterminacy and Arbitrariness A more general substantive problem with the Duty View is the degree of indeterminacy and arbitrariness that comes with the proliferation of a large number of duties and rights, each with uncertain borders and relations to the others. Among those that play a role in Tadros' argument are a wrongdoer's duties to thwart her crime, to provide a remedy, to pool her protective duties with those of other wrongdoers, and to submit to deterrent punishment; a victim's duties to punish wrongdoers and transfer his right to punish to the state; and the state's duties to protect citizens from crime and from unjust punishment. According to Tadros, some of these duties may fade over time as the duty-holder becomes less psychologically connected to the person he was when he committed the crime.

Needless to say, with duties as inherently broad, vague, and temporally unstable as many of these are, problems of interpretation and application are daunting. For example, when is the remedial duty satisfied and the debt paid? How should we measure the state's duty to punish in proportion to the good that would accrue -- case-by-case, or systemically with all

punishments treated collectively? If the former, how would we isolate the effect of the individual's punishment? As duties and rights proliferate, and as more than one applies (or is available to be applied) to a particular circumstance, attempts to define, apply and balance them will generate a large margin of error that, with successive iterations, threatens to take over the page. As I shall now argue, we get closer to justice with a discourse that is tied more directly to our moral intuitions and capable of finer distinctions than the abstract discourse of rights and duties that constitutes the Duty View.

2. A METHODOLOGY OF DIMINISHING RETURNS

Let us now put aside these substantive critiques, assume that all of Tadros' subsidiary claims are individually plausible, and consider how they operate collectively in an argumentative chain. I want to examine two inherent limits on the persuasiveness of highly distended moral reasoning of this type. The first problem is that, as a general rule, the more complex and lengthy the argumentative chain, the less confidence we should have in its conclusion. A chain of inductive reasoning is weaker than the sum of its parts. The second problem is comparative. When such an extended chain of reasoning is necessary to establish a position as revisionist as the Duty View, we may think that it isn't enough to justify rejecting much more deep-seated and immediate intuitive beliefs. However strong the argument, it will lack plausibility if it is incompatible with fundamental moral convictions that are too compelling to doubt. (I leave aside a third methodological critique that has been persuasively demonstrated elsewhere: Tadros' reliance on highly idiosyncratic hypotheticals to elicit far broader principles than they can support. See Husak 2012: 19)

Consider first Tadros' argument on its own terms. Its initial steps invoke certain intuitively plausible principles, such as a moral duty of non-aggression; succeeding steps are mainly established inductively by taking a preceding step's principle and eliciting responses to hypotheticals testing its extension. This multiple-step moral argument is essential because the ultimate principle it seeks to establish -- that state punishment is permitted only insofar as it may fulfill the wrongdoer's protective duty -- is not at all intuitive by itself.

This kind of moral reasoning suggests the construction of a building, starting with the foundation and progressing upward as each successive floor is built. It appears to make progress by addition. But addition is a misleading metaphor for viewing this kind of argument. Each successive step should reduce our confidence in the conclusion,[8] for three reasons:

8 Unless, of course, it adds to the plausibility of a prior one.

(i) The longer the chain, the greater the chance for a substantive error to infect it. Each additional step brings with it an additional risk of failure.

(ii) The second reason applies even if every step is highly likely to be correct. It reflects the mathematical truth that a chain of inductive reasoning is not "as strong as its weakest link" but weaker. The "weakest link" adage *does* apply to deductive reasoning, where the truth of the logically-entailed conclusion rests entirely on the truth of its premise. But Tadros' argumentative steps are based on inductive inference, analogy, and intuition, none of which can supply the 100% confidence that logical entailment does. In this case, each step can only be judged more or less plausible, and each step makes the chain weaker by compounding the possibility of error, however minimal. A Bayesian calculation would treat each step as probabilistic to the degree of its plausibility, and the likelihood of the concluding proposition as a product of the multiplication of fractions, just as the chances of tossing two tails in a row are ½ x ½ = ¼.

We can illustrate how severely Tadros' methodology undermines his thesis by assuming that each individual step in Tadros' argument is amply persuasive, with all intuitive and inferential claims highly plausible. If we represent this arithmetically by assuming a 90% level of confidence in each of the eleven steps, the likelihood of the conclusion being correct is just 31%. We may well have more confidence in the brute conviction that only desert can justify and calibrate punishment.

Now this 31% figure is illustrative and subject to reasonable disagreement. One might deem some step unnecessary, or deductive, or so self-evident as to be incontrovertible, for example. But at least five of the eleven steps would have to be entirely discounted on such grounds to render Tadros' conclusion even slightly more likely than not.

(iii) Some may question whether this mathematical likelihood of error is sufficient to discredit a moral argument in the absence of specific counterarguments, or whether mathematical probabilities can be attached to moral judgments at all. But a third reason to doubt Tadros' distended argument needs no mathematical proof to warrant acceptance: *Occam's Razor,* the principle that the simpler explanation for a proposition is more likely to be true than a complex one, all else equal. Given its fruitfulness in directing scientific investigation over centuries, we need not be able to

explain *why* this principle is true to be justified in presuming it is.[9] By this standard, Tadros' 11-step argument should be rejected as unnecessarily complicated, if only because the same principle Tadros uses to justify the victim's right to self-defense – the Choice Principle – would justify state punishment directly if punishment deters crime. If imposing the death penalty on convicted killer A will deter the unprovoked murder of B, the Choice Principle provides a reason to execute A because, as one of the two will die, it is fairer that it be the one who had the opportunity to avoid the risk.

In his response to my "diminishing returns" argument, Tadros does not contest the math, or the inverse relation between the number of steps he uses and the likelihood his conclusion is correct. He argues instead that all moral claims rely on the truth of many subsidiary claims, so all moral claims are vulnerable on this analysis, including the ones I propound elsewhere in this essay. My "diminishing returns" argument would lead to a general moral skepticism, he claims. (2015b: 58)

But that's the wrong lesson. My argument is not an invitation to moral skepticism, nor a claim that intuitions are infallible, but an appeal to consider the relative persuasiveness of different legitimate modes of moral discernment on the question at hand. As moral argument becomes increasingly abstract and distended, the intuitive plausibility of its conclusion increases in importance. This creates a burden that highly revisionist moral arguments may not be able to meet. The problem with Tadros' multi-step argument is that it culminates in a rationale that opposes the fundamental intuitive convictions most people have about the morality of punishment, and this forces us to decide which ground of belief is more trustworthy on the issue. This is one answer to Tadros' claim that my arguments are as vulnerable as his own because they also involve multiple steps. The conclusions to those arguments - that it is unjust to inflict punishment in the absence of desert, and that it is unjust to the victim for the state to ignore his victimization – do not require us to choose between an extended chain of reasoning and our considered moral convictions.

The bottom line is that Tadros' methodology places his conclusion in a less plausible initial position before any question is raised about the soundness of individual links in the chain. This doesn't obviate exploring the merits of those links, see Sec. 1(B) above, but it does provide a substantial, independent ground for skepticism.

9 Richard Swinburne (1997) claims that "it is an ultimate a priori epistemic principle that simplicity is evidence for truth." Some argue that this principle is self-evident, constitutive of rationality, or another kind of foundational truth that cannot be further justified. Others, however, accept Occam's Razor is an appropriate methodological maxim but not by itself indicative of truth.

3. COUNTER-INTUITIVE SENTENCES

The last section examined the intricate structure of Tadros' reasoning and whether it can support the principles he derives from it. If we examine how these principles would apply to specific cases, the difficulty multiplies. They produce results so counter-intuitive that something has to give. Apparently recognizing this, Tadros modifies the Duty View to make it cohere with moral common sense, but to such an extent that it largely disappears; as I shall argue, with Tadros' many work-arounds in place, what remains is something very much resembling negative retributivism. This is not surprising, as so many obvious sentencing factors are easily understood in terms of retributive desert but invisible to the Duty View in its unvarnished form.

3.1 Losing Proportionality in Punishment

Consider the case of Dzhokhar Tsarnaev, recently convicted for his part in the 2013 Boston Marathon bombings that killed three people and injured hundreds of others. The Duty View would prohibit punishing Tsarnaev if he were innocent, but is it capable of generating a *proportional* sentence to Tsarnaev given his conviction? I doubt it, unless we so revise our moral convictions as to make irrelevant numerous factors that Anglo-American jurisprudence has consistently viewed as important considerations in sentencing.

For example, in planting the bombs Tsarnaev acted with extreme premeditation. Others may kill on impulse, or after being provoked, or by negligence, or by accident. Retributivism can explain the enormously different sentences imposed in killings perpetrated with these different levels of culpability: a premeditated killing displays a degree of blameworthiness that doesn't exist in a negligent one, for example, and therefore deserves far harsher punishment. On the other hand, Tsarnaev was a teenager, and we are likely to think he therefore deserves a different, lesser sentence than a forty year old career criminal who committed the same crime. These factors have been important to sentencing judges because of their obvious relation to desert. Retributive sentences are straightforward in that way; most retributivists believe that a punishment should reflect the gravity of the crime and the blameworthiness of the perpetrator in committing it. If premeditation, youth, and prior convictions are relevant sentencing considerations under the Duty View at all, there is nothing straightforward about why this is so.

On the Duty View, we impose sentences based on many factors, but most centrally on the factors that governed the wrongdoer's liability to be harmed defensively at the time of his crime. The two primary sentencing limitations are that the harm it inflicts on the wrongdoer not exceed that

which (A) he would have been liable to suffer from the victim defending himself (2011: 347), *and* (B) is warranted for the sake of the net benefit it will produce (333-34). That sets the baseline for the wrongdoer's liability to suffer harm in punishment, which is then adjusted by many factors, including *inter alia* reductions based on repentance (347-48), the passage of time since the crime (347-48), the difference between "eliminative" self-defense and "manipulative" punishment for deterrence (319-320), and the offender's absence of responsibility for the threats his punishment will deter (348).

As to "A", note that this proportionality requirement has little to do with the rectificatory basis that Tadros invokes as the primary ground for liability. The sentence is limited not by what would be required to make the victim whole, nor by what would now be required to avert a new but similar crime against him, but by how much harm the victim was permitted to inflict defensively at the time of the crime. This seems to make the aggressor's liability to punishment dependent on the factors that govern the permissibility of self-defense. Key among them are necessity and proportionality: the defender may use only the amount of defensive force that is necessary to repel the crime and proportional to the gravity of the threat it is defending against. There are other factors that may limit defensive force, but none of them are sensitive to the age, record, or culpability of the aggressor because those factors do not change the moral preference afforded the victim given that one of the two must suffer harm. Nor do age, record and culpability have much bearing on compensation to someone victimized by a crime; whether the aggressor's threat is a product of negligence, recklessness, or design is not at issue, only the degree of harm that must be rectified.

On the other hand, some factors that *are* relevant to self-defense have little obvious bearing on punishment. Most inapt is the self-defense element of necessity. A victim defending himself may use only the amount of force necessary to repel the aggressor, which means that differences in the circumstances – such as the type of weapon the aggressor is using -- will change the amount of defensive force permitted. If one's liability to punishment depends on one's liability to suffer defensive force, punishment may vary greatly among wrongdoers who are identical in every respect except the morally arbitrary circumstances in which they acted. The result will be sentences that, intuitively, seem much too low or much too harsh. Consider as examples:

- A, a bank teller, embezzles V's account. B, a purse snatcher, grabs V's purse. Should A and B's liability to punishment vary according to the fact that V is entitled to use defensive force against B but not A?

- A shoots V and misses, leaving V unaware of his narrow escape. V has no need to defend himself and no right to compensation. Does this have any bearing on whether A should be liable to punishment for attempted murder?

- A threatens V with a knife. V is able to retreat safely and therefore defensive force is unnecessary; but V fails to retreat, parries A's thrust unsuccessfully, and dies. Does the fact that A was not liable to be harmed in self-defense have any bearing on what his liability to punishment should be? In his response to this last hypothetical, Tadros seems to argue that the permissibility of defensive harm does *not* have bearing in that case, but that seems impossible to reconcile with his fundamental argument against desert.[10]

3.2 Tadros' Work-arounds

This mismatch between permissible self-defense and proportional punishment should not be surprising. Even if we assume the purpose of both is to prevent crime, the factors relevant to averting a wrongdoer's imminent threat are not the same as those relevant to using a wrongdoer as a means of preventing crimes by other people against other victims at some time in the future. Tadros allows that the "transition from self-defence to punishment may not be entirely smooth," (2011: 348) and develops a number of arguments for departing from the self-defense template he had adopted. The challenge he faces is to change the scope of punishment liability -- expunging the irrelevant factors and incorporating the relevant ones -- while somehow maintaining the self-defense rationale for liability. In my judgment it is a challenge that defeats him: the independent proportionality rationales he marshals become so ad hoc, and so divorced from the self-defense and compensatory grounds that preceded them, that we may wonder what the self-defense template for punishment has accomplished. Some illustrations:

Punishing environmental crimes - Most criminal codes include environmental crimes that cause harm without harming any particular person – for example, the intentional killing of the last member of a marine species. Duties of victim compensation can't apply to such a case, so one can't justify punishment

10 Tadros says that V wrongs A because his defensive force was unnecessary, but that "given his wrongful act he is permitted to defend himself. It follows that there is no asymmetry between liability in self-defence and liability in punishment of the kind that Blumenson's argument relies on." (2015b: 71). This seems to contradict the central argument Tadros used to show that liability to defensive force is justified by the Choice Principle, not by the aggressor's desert. That principle of distributive justice holds that if harm is inevitable, it is better that it befall someone who was responsible for creating the situation than one who is not. In this case, harm is not inevitable given the victim's opportunity to retreat; and when he fails to do so, he is the person responsible for making harm to one of the two inevitable.

by extending the compensatory duty. But Tadros doesn't interpret that fact to bar punishment of such crimes. Rather, he develops some much more direct grounds for punishing them.

On Tadros' alternative rationale, the wrongdoer had a duty to respect the natural environment, and his violation of that duty may give rise to "a duty to protect that aspect of the natural environment from further damage. If the first duty is not owed to anyone, neither is the second…. Hence, it may be permissible to punish [him] to deter others from causing further damage to the natural environment." But the duty not to commit a criminal act applies to *all* crimes, raising the question of why these grounds don't obviate his more complex (and therefore, *ceteris paribus,* less persuasive) primary argument. Why isn't that duty sufficient to ground a protective duty to prevent further such crimes, without relying on a particularized duty to victims at all? Richard Burgh (1997: 316) has made an argument along these lines, characterizing a crime as a social harm that requires the offender to compensate society through punishment.

Punishing attempts - Similarly, an attempted crime does not harm a particular victim who has no knowledge of it. If the perpetrator of such an attempt is to be punished, it can't be because of any liability to submit to defensive force or provide compensation to the victim. However, Tadros thinks these crimes *can* be punished because attempted crimes divert police resources and make *everyone* less safe, thereby establishing the attempter's duty to compensate everyone (2011: 326-27). Alexander (2013) has criticized the factual premise of this claim,[11] and Ferzan (2013: 185-86) has criticized its rationale, both because it offers no way to determine what degree of harm a wrongdoer's diversionary act has caused, and because it holds her strictly liable for it. But even if Tadros could answer these critiques, his rationale would still dictate an entirely implausible proportionality calculus. For example, it suggests that attempted drug smuggling should be punished more harshly than attempted murder, given the greater police resources devoted to preventing smuggling.

Since publication, Tadros has turned to a different argument to justify punishing attempted crimes. He argues that it is implausible that "attempting to harm others makes no difference to a person's liability to be harmed as a means…I do not see what argument could be provided for it. Even if we think that causing a threat is very important to ground a person's liability to avert the threat, why should we conclude that attempting to cause harm

11 Alexander (2013: 172) argues that because many failed attempts would go unnoticed but for their criminality, it is "ludicrous to assert that were they not criminalized, they would be causing us to devote security resources away from averting harmful acts."

is insufficient on its own to *make any difference at all* to a person's liability to be harmed?" (2013: 320). Tadros' intuition is clearly explainable on grounds of desert, but given his rejection of that ground, and the inapplicability of Tadros' compensatory theory, I would argue this gets the burden of persuasion backwards, that the argument missing is his own.

Punishment beyond compensation - Attempted crimes present one of several difficulties that confront Tadros' effort to justify punishment on the basis of the offender's residual compensatory obligations. As commentators have noted, this basis also produces the unacceptable corollary that wealthy offenders may be able to buy their way out of punishment (Ferzan 2013: 189-91; Walen 2012). Most people would find this unacceptable on grounds of inequality, but because their measure of inequality is comparative desert, that can't be Tadros' worry. His worry is that an offender who does not bear any significant cost in protecting the victim "cannot claim to have fulfilled his rectificatory duty.…[f]or he would only have done what he would have had a duty to do independently of his wrongdoing" – the duty to rescue each other from harm if it can be done at little cost (2011: 286). So Tadros annexes an independent basis of punishment liability to the Duty View. He claims that:

> our compensatory duties are fulfilled when we have done what is required to rectify the harm that we have caused. The duties that underpin punishment, in contrast, are not…Because I could have been harmed as a means to avert a threat that I posed as a result of my wrongdoing, I may now be harmed as a means to an equivalent degree to avert other threats, even threats of a greater magnitude than the harm that I caused (288, 291; also see 283-91 discussing punishment beyond rectification).

The question is, *Why?* Tadros offers two rather vague answers. The first is the analogical argument that because a guarantee of compensation does not obviate the prohibition on tortious conduct in advance, *ex post* compensation does not fulfill the offender's duties either (2015a: 82-83). But this analogy works only if the reason compensation is insufficient *ex ante* also applies *ex post,*

and it doesn't. That reason depends on the act not having occurred: at that point, compensation is second best to preventing the tort to begin with. *Ex post* no such preference exists. In fact, there can't be a preference as to type of remedy, because both the compensatory and further duties are paid in the same currency - protection from future crime.

The best we get is the discussion of a hypothetical -- *Three Threats,* described below[12] - that suggests that committing a crime subjects the wrongdoer to conscription as a utilitarian means for crime control. Tadros' grounds are that (1) the victim is in serious danger, (2) the offender is harmed to no greater degree than he is liable to be harmed to avert the threat he imposed, and (3) he could have avoided that liability simply by refraining from his wrongful act (2013: 303, enunciating these reasons for the Three Threats conclusion; 2011: 291). But these grounds prove too much twice over. The first difficulty is that these grounds cannot distinguish between using wrongdoers to prevent crime and using them to supply organs or fulfill other important social needs. In order to avoid making the offender fair game generally, Tadros offers some exceptions to the above rationale; whether these exceptions are persuasive I leave to the reader to consider.[13]

12 In Three Threats, Tadros imagines that Bob has propelled a boulder that will injure Jane. On the Duty View, Jane would have been permitted to harm Bob to Y degree in order to stop the boulder, but the boulder is unstoppable. She can, however, use Bob to divert either Boulder 2 or Boulder 3, each of which have been pushed towards her by others. If she uses Bob to avert Boulder 2, she will avert the same degree of harm that Bob's boulder will cause, at minimal injury to Bob. If she uses him to avert Boulder 3, she will avert twice the degree of injury that Bob's boulder will cause, but Bob will be much more badly injured, though in an amount less than Y (2011: 289). Tadros argues that it is strongly intuitive that Jane may use Bob to avert Boulder 3, even though using him to avert Boulder 2 would fully satisfy Bob's compensatory obligation with less injury to him (291).

13 Tadros says that certain kinds of punishment will always exceed the offender's liability. On his account, even if an offender was subject to lethal defensive force at the time of the crime, capital punishment is impermissible given the passage of time and the difference between eliminative and manipulative harm (2013: 308). What about the non-lethal harm of organ removal for transplantation? Tadros rules this out as well on the following grounds: (1) "It is wrong to harm a person to tackle a problem that is utterly different from criminal offending." (2011: 354). Whether a particular punishment is excluded thus depends on what kind of differences matter and why, but we get no more than the conclusory term "different." Without a definition, the criterion cuts both ways: others would find the difference between (a) harming a wrongdoer to defend against his aggression and (b) harming him to deter the aggression of others sometime in the future sufficient to make the latter impermissible. (See Quinn (1985). (2) "It may be that there is something special about organ distribution…. Perhaps using a person's organs is a particularly pernicious way of using a person…" (2013: 309). But again, without specifics we are hard-pressed to explain why incarcerating a person regardless of his desert is not similarly pernicious. (3) Institutional reasons to constrain punishments: for example, the likelihood that the institutions administering transplant punishments will act unfairly, the availability of less draconian but equally effective alternatives, and the expansive tendencies of criminal justice systems (2013: 308-09). Notably, however, all three grounds may easily apply to exclude prison sentences as well.

The other way that Tadros' reasoning proves too much is that it seems to leave the duty of compensation superfluous. Its justification reaches all the punishments that were previously grounded in the compensatory duty. The latter now seems beside the point - as arguably it should be, given that a victim may be fully compensated through the civil system.

Let us consider a different explanation for the intuition that wrongdoers have an extra-compensatory duty to deter crime by their own punishment. Daniel Farrell (2015) has argued that this putative requirement must be a duty of retributive justice; Tadros disputes this on the ground that "retributivist views of punishment are not grounded in duties of offenders, or the relevant duties have nothing to do with protection." (2015a: 81)

I believe Tadros is right to deny that this additional duty constitutes the pure form of retributivism that requires the infliction of suffering on offenders, because his theory prohibits doing so solely for that purpose. But I find it difficult to view his claim as plausible except as an implicit if diluted version of the negative retributivist theory I argue for in the following section – the view that desert is a necessary but not sufficient condition for punishment. On the Duty View, the offender has a duty to exceed compensation to the full extent of his liability to be harmed, but this duty should only be enforced if some utilitarian benefit will accrue. And on what basis does that duty exist? Not on the basis of either utility or compensation, because the duty exists apart from both. In the absence of some other specification by the author, it seems that the duty persists beyond full compensation because the offender deserves to suffer the additional burden. That seems to be the best explanation for Tadros' statement that assailant Jake still owes something after he has fully compensated his victim Sally: "If Jake manages to benefit Sally at little cost to himself, we will have a sense that he has 'got away' with his crime." (2011: 289).

4. THE INEXORABLE SIGNIFICANCE OF DESERT

I have just argued that the Duty View generates highly counter-intuitive instructions to a sentencing authority unless modified beyond recognition, and that retributivism offers a better account of the factors that must inform the proportionality calculation. This should count heavily against the theory *unless* its retributivist rival is itself lacking in some greater way. So we must assess the comparative strength of the retributivist alternative and Tadros' arguments against it.

Tadros argues that retributivism is not an option because it is both false and incoherent, and what intuitions seem to support it can be better explained

in other ways. His central objection, and the one I shall explore here, is that retributivism rests on a false premise – the premise that a wrongdoer's suffering is intrinsically valuable to the degree it is deserved (2011: 26, 45).[14] But there are many retributivisms, and only some of them depend on that premise.[15] Therefore Tadros' critique cannot establish the Duty View's comparative advantage over retributivism *tout court*.

More specifically, Tadros' definition does not encompass retributivists who believe that punishing criminals is at least sometimes a duty of justice, a claim I shall argue in part 4(B). Nor can it apply to the prominent version known as negative retributivism. Negative retributivists do not believe that the state *should* punish a criminal to the extent she deserves. Rather, they believe that the state *must not* punish the innocent, nor punish the guilty beyond what they deserve. For them, desert is a necessary but not sufficient condition for punishment; therefore punishment is justified only if there is also an additional, non-retributive basis.

Negative retributivists quite clearly do not believe that the "goodness" of deserved suffering requires the punishment of a criminal, and there is no reason to think they believe such suffering is good at all. *But desert is still the central element in their theory:* its absence bars punishment, and its presence places an upper limit on the permissible sentence. Negative retributivism illustrates the error in Tadros' argument: one can't eliminate the importance of desert to sentencing by showing that deserved suffering is, like all suffering, bad. We can believe this but also believe that the infliction of suffering in proportion to desert is sometimes justified as a necessary evil.

There is a reason Tadros dismisses negative retributivism, but not, I think, a good one: he believes there can be no such thing. He claims that by its terms, retributivism provides a putative reason to punish the guilty but no reason at all against punishing the innocent, or against punishing the guilty beyond what they deserve. He further argues that if it did protect the innocent, the idea would be incoherent. The next section challenges these two claims. Following that, section 4(B) shows how a broader non-consequentialist theory of punishment might justify both a limit on punishment (through negative retributivism) and an affirmative duty to punish (grounded on other considerations of justice), as one example of a theory incorporating desert that is not subject to Tadros' critiques. I suggest that this theory is more compelling than the Duty View, given its superior ability to account for our considered sentencing judgments and its greater

14 More precisely, Tadros believes that only a retributivism built on that premise would be weighty enough to justify an incarcerative system of punishment.

15 Those that do include the theories argued in Moore (1997) and Kirshnar (2000).

coherence as a theory.

4.1 Desert as Prerequisite: Is Negative Retributivism Possible?

Retributivists disagree about whether their theory commands punishment of the guilty or only permits it, but it seems *all* retributivists believe that their theory bars punishment of the innocent, and consider this a unique and powerful point in their favor. Remarkably, this is precisely the opposite of Tadros' novel rendition of retributivism, which to him constitutes only a sword, not a shield. This follows, he says from "the very simple truth that the existence of a reason to do something does nothing to exclude the possibility that the thing (and more) could not adequately be justified on other grounds." (2011: 36; also see 35-7, 312-13). He claims that only the Duty View protects the innocent from punishment.

The oddity is that the basis for Tadros' claim that the Duty View bars punishment of the innocent is identical to the reason retributivism purports to do so. That reason is the Means Principle. The claim in both cases is that only the guilty may be punished consistent with the Means Principle - either because only they deserve it (according to retributivists) or because only they have a protective obligation the state may enforce (according to Tadros). Given the parallel structure, there are no good grounds to claim, as Tadros does, that the Duty View bars punishment of the innocent but retributivism does not.

Perhaps Tadros believes that the Duty View includes the Means Principle, while the retributive view does not. But why should he think that? The retributivist tradition since Kant has most centrally embodied respect for the right of autonomous individuals to choose their ends for themselves, a constraint that restricts punishment to those who will it on themselves by their own blameworthy acts.

Tadros (2015b: 59) also argues that if negative retributivism did protect the innocent it would be incoherent because desert necessarily cuts both ways:

> I think that 'X deserves O' implies that there is good reason for X to get O irrespective of any further good that will be secured if X gets O. If so, there is no such thing as negative desert in Blumenson's sense.

A negative retributivist may believe it bad (or unjust) for someone to get what he doesn't deserve without being logically committed to the view that it is necessarily good (or just) for him to get what he does deserve (Wasserstrom 1978: 309-10). But let us assume that Tadros' first proposition is correct. His second sentence would not follow from it if any one of the following is true:

I. A reason may be overridden. An offender may deserve punishment, and this may entail that punishing him is good in one way, but

the net costs of punishment may be so great as to foreclose it. Here's an analogy: equality is intrinsically good in one way; if achieving it would require redistribution that left everyone almost as impoverished as the worst off were, doing so may be bad in a greater way; therefore, whether states should aim for equality in any particular case is contingent on the costs and benefits it would produce.

II. A moral side-constraint might bar the state from aiming to cause offender suffering even when deserved. That constraint would not rule out state punishment for other reasons.

III. Even if "X deserves O" entails that X should receive O, desert will not require punishment if "O" signifies "forfeiture of a liberty right" rather than "punishment." The question here is whether one can deserve to lose a right by his blameworthy actions, and surely it is at least coherent to think one can. Some of one's rights may be contingent on respecting the rights of others. Alternatively, blameworthiness may be the fairest criterion for the distribution of individual punishments when a punishment regime will produce enough benefit to be justifiable. There is nothing unintelligible or inconsistent in recognizing both a moral prohibition on punishing the blameless and a humane directive to impose punishment on the guilty only when something would be accomplished by doing so.

Rights-forfeiture theories differ over what rights a criminal forfeits by her conduct. Depending on the theory, she might retain rights against punishments that serve no utilitarian benefit, or inflict pain, or are imposed by vigilantes, for example (see Wellman 2012). There are many versions, but to qualify as a form of negative retributivism, the forfeited right against punishment must be limited to punishments proportional to the offender's desert.

4.2 Beyond Desert: Why Punish?

The second and third interpretations conceive negative retributivism as a matter of principle – one that that limits who may be punished and by how much – but by itself provides no reason to punish anyone. If there is to be any punishment at all, negative retributivists must look to some other theory that provides an affirmative reason to impose punishment on an offender. Many negative retributivists find that reason in the utility of punishment as a crime-preventive deterrent. That hybrid view parallels Tadros', except that the permissibility of punishment is grounded in desert rather than a forward-looking remedial duty.

But there is powerful intuitive support for another reason to punish as well: as Tadros notes, most people believe that "something is amiss when a serious wrongdoer is not punished," even in the absence of deterrent value (2011: 276). In this final inquiry, I add this conviction to the mix and ask whether it weighs for or against either theory.

If we credit this intuitive conviction as roughly reflecting some moral truth, we must ask what exactly *is* amiss. Tadros thinks it is the fact that "offenders who are not punished have a duty that is unfulfilled." (276). The retributivists he targets believe it is the failure to inflict suffering on wrongdoers who deserve it. One might argue that a negative retributivist who rejects both views, as I have, can't explain this intuition and that this should count heavily against his theory. There are two rejoinders. First, even if true, that criticism would not count against the theory in a pair-wise comparison with the Duty View, which itself leaves a wrongdoer unpunished when deterrent value is lacking. Second and more importantly, a negative retributivist *can* answer that challenge because, unlike Tadros' theory, nothing prevents him from combining a desert-based limit on punishment with a non-utilitarian rationale for imposing it.

Here is a rudimentary sketch of one possibility: a hybrid theory that asserts as state deontological obligations (1) a prohibition on inflicting undeserved punishment, and (2) a responsibility within that limit to impose punishment when required as a matter of justice to the victim. (Whether utilitarian benefit is also a reason to impose a permissible punishment is a separate question.) The intuitive appeal of the second element is currently illustrated by the demands for "justice for Michael Brown" by residents of Ferguson, Missouri, following a grand jury's failure to indict the police officer who killed him, and by the startling number of similar cases since. What is "amiss" in such cases, and in a state's refusal to sanction any grave crime, is the injustice that inflicts upon the victim: the devaluation for a second time of someone who has already been treated by the perpetrator as no more than an instrument to his ends. For George Fletcher, this constitutes state complicity in the crime that leaves "the victim's blood…on our hands." (Fletcher 1995: 6, 205; also see Hampton 1992: 1684, 1692; Burgh 1987). Complicity may be too strong, but at the least, such state inaction betrays the protective role that largely underwrites its own legitimacy, and treats the victim as an outcast.

On this account, the state's obligation to punish crime derives from the injustice it does to victims when nothing is done. But to be clear, this rationale for punishment is a qualified one: punishment is not the obligation, only a means of fulfilling it, and whether other means may also do so is necessarily

dependent, at least in part, on the social meanings that prevail in a culture.[16] In ours, many people believe that a long term of imprisonment is the only way of taking victimization seriously. Yet it is possible to imagine a cultural shift towards less draconian methods, such as fines, community service, house arrest now enforceable through the use of GPS ankle bracelets -- and even non-criminal restorative justice resolutions, in which case this justification for punishment would dissolve.

To be sure, the Duty View is also concerned with justice to the victim -- unlike the prevailing punishment theories, which neglect the victim as an independent subject of justice[17] -- but its conception of justice is quite different, and is contingent on its crime-prevention utility. Tadros believes that the state's obligation to repudiate crime and vindicate its victims can't justify punishment, (2011: 87-8, 91-2, 107, 109; 2013: 255) and argues that because a victim's moral status "is incapable of being eroded through wrongful action….it is difficult to see why the obligation to protect people against lack of respect is very significant in itself" (2011: 108).[18] But this misses the real stakes involved in a state's response to victimization, and leads Tadros to severely underestimate its importance. What is involved is not merely the wrongdoer's violation of a victim's inalienable moral status, but the very *existence* of the victim's civil status as a member of the political community. In the United States, it is unconstitutionally cruel and unusual punishment to strip a person of "his individual status in organized society".[19] A state that ignores crimes against its citizens withdraws that status.

There are other hybrid theories, of course. H. L. A. Hart's distinction between the utilitarian "general justifying aim" of punishment and a desert-based distributional constraint is perhaps the most influential among many (Hart, 1968: 8, 10). The only point here is to demonstrate the inexorable significance of desert in punishment as a restriction on its permissibility, and its compatibility with a range of affirmative rationales

16 A further, necessary question, of course, is whether any particular cultural take on the requirements of justice may be morally unacceptable. Where permissible cultural diversity ends and moral imperatives begin is a vexing problem, and one that increasingly confronts institutions of global justice. For the International Criminal Court, that issue takes legal form as the question whether non-criminal, transitional justice alternatives are sufficient to bar ICC prosecution under its complementarity principle. Rome Statute of the ICC, Art. 17, UN Doc. A/CONFR. 183/9 (July 17, 1998). Over time, ICC decisions on the issue will define minimum requirements for criminal justice systems throughout the world.

17 Consequentialism is concerned with the collective benefit, presumptively leaving the victim with no greater significance than any other individual; retributivism classically treats justice to the victim as if it were a mere by-product of, or necessary identical to, the punishment that is required to treat the offender as she deserves.

18 See also 2013: 277-79, in which Tadros rejects Adil Ahmad Haque's claim that punishment can be justified because we owe it to the victims; see Haque 2013.

19 Trop v. Dulles, 356 US 86, 101-102 (1958).

for punishment. What's fatally counter-intuitive in the Duty View is its exclusion of any direct role for desert as a prerequisite to the permissibility of punishment.

CONCLUSION

I have argued that the Duty View is unpersuasive as a theory of punishment, and that an alternative theory – one that recognizes both a defendant's right against undeserved punishment and a victim's right to vindication – comes closer to a point of reflective equilibrium in which principles, theories, and intuitive judgments cohere and support each other rather than forcing us to choose among them. Yet what a fruitful thing it is for the philosophy of punishment that Tadros has made his case for the Duty View! There are not many works within the field that cover so much criminal law ground with such originality, and even fewer with the potential to stimulate a new wave of thought on numerous issues in the field.

And, it is to be hoped, not just in the realm of philosophy. One of the gifts of a theory of punishment as imaginative and revisionist as the Duty View is that it allows us to view our correctional policies through a new lens, rethink old choices, and discover alternative routes to security that don't always go through a prison gate. *That* route, which the United States has traveled for the last half century, has left us with over 2.2 million prisoners and the distinction among nations of imprisoning the largest percentage of its population, five times the world's average (Bureau of Justice Statistics 2012: 3). Tadros asks us why, and what we have to show for it. His approach presses us to take the suffering of inmates seriously as a central moral element in punishment, and in so doing, to consider alternatives. For consequentialists, retributivists, policy-makers, judges and others who long for a morally defensible criminal justice system, that is the best place to start.

BIBLIOGRAPHY

Alexander, L., 2013: "Can self defense justify punishment?", *Law and Philosophy* 32: 159-75.

Brun, G., 2014: "Reflective equilibrium without intuitions?", *Ethical Theory and Moral Practice* 17: 237-52.

Bureau of Justice Statistics 2012: "Correctional Population in the United States, 2011", in Bulletin NCJ 239972 : 3.7-52.

Burgh, R., 1987: "Guilt, punishment, and desert", in *Responsibility, character, and the emotions: new essays in moral responsibility*, ed. F. Schoeman, 316-37, Cambridge: Cambridge University Press.

Duff, RA., 2013: "Punishment and the duty of offenders", *Law and Philosophy* 32: 109-27.
Farrell, D.M., 2015: "Using wrongdoers rightly: Tadros on the justification of general deterrence", *Criminal Law and Philosophy*, 9: 1-20.
Ferzan, K.K., 2013: "Rethinking the ends of harm", *Law and Philosophy* 32: 177-98.
Fletcher, G., 1995: *With justice for some: victims' rights in criminal trials*, New York: Addison-Wesley Publishing Company.
Hampton, J., 1992: "Correcting harms versus righting wrongs: the goal of retribution", *UCLA Law Review* 39: 1659-1702.
Haque, A.A., "Retributivism: the right and the good", *Law and Philosophy* 32: 59-82.
Hart, H.L.A., 1968: "Prolegomenon to the principles of punishment 1", in *Punishment and responsibility*, 1-27, Oxford: Clarendon Press.
Husak, D., 2012: "Retributivism in extremis", *Law and Philosophy* 32: 3-31.
Kirshnar, S., 2000: "A defense of retributivism", *International Journal of Applied Philosophy* 14: 97-117.
Moore, M.S., 1997: *Placing blame: a theory of the criminal law publisher*, Oxford: Oxford University Press.
Nozick, R., 1974: *Anarchy, state and utopia*, New York: Basic Books.
Quinn, W., 1985: "The right to threaten and the right to punish", *Philosophy & Public Affairs* 14: 327-73.
Rawls, J., 1971: *Theory of justice*, Cambridge, MA: Belknap Pess of Harvard University Press.
—1999: "The independence of moral theory", in Rawls, J., *Collected papers*, ed. Samuel Freeman, 286-302, Cambridge, MA: Harvard University Press.
Swinburne, R., 1997: *Simplicity as evidence for truth*, Milwaukee, WI: Marquette University Press.
Tadros, V., 2011: *The ends of harm: the moral foundations of criminal law*, Oxford: Oxford University Press.
— 2012: "Replies", *Jerusalem Review of Legal Studies* 5: 89-109.
— 2013: "Responses", *Law and Philosophy* 32: 241-325.
— 2015a: "Answers", *Criminal Law and Philosophy* 9: 73-102.
— 2015b: "Response to Blumenson", *Law, Ethics and Philosophy* 3: __.
Tanguay-Renaud, F., 2013: "Victor's justice: the next best moral theory of criminal punishment?", *Law and Philosophy* 32: 129-57.
Walen, A., 2012: Review of the book *The ends of harm* by Victor Tadros, at Rutgers University, *Criminal Law and Criminal Justice Books*, http://clcjbooks.rutgers.edu/books/ends_of_harm.html
Wasserstrom, R., 1978: "Some problems in the definition and justification of punishment", in *Values and morals*, ed. A. I. Goldman and J. Kim, 299-315, Dordrecht: D. Reidel.
Wellman, C., 2012: "The rights forfeiture theory of punishment", *Ethics* 122: 371-93.