Law and Philosophy (2013) 32: 159–175
DOI 10.1007/s10982-012-9157-y

LARRY ALEXANDER

CAN SELF-DEFENSE JUSTIFY PUNISHMENT?

(Accepted 24 September 2012)

ABSTRACT. This piece is a review essay on Victor Tadros's The Ends of Harm. Tadros rejects retributive desert but believes punishment can be justified instrumentally without succumbing to the problems of thoroughgoing consequentialism and endorsing using people as means. He believes he can achieve these results through extension of the right of self-defense. I argue that Tadros fails in this endeavor: he has a defective account of the means principle; his rejection of desert leads to gross mismatches of punishment and culpability; and he cannot account for punishment of inchoate crimes.

Can one advocate a punishment system based on deterrence without being a thoroughgoing consequentialist and with an objection to 'using' people as means to produce others' well-being? Easy as pie. One can make the defendant's desert a limitation – a ceiling – on the severity of his punishment, but make any punishment within that limitation deterrence-based.

However, what if one is also a desert-skeptic? Suppose, that is, that one cannot endorse the desert limitation because one does not believe in desert? This is the challenge Victor Tadros sets for himself in his new book, *The Ends of Harm*.[1]

Tadros sets out to show that one can justify an attractive instrumentalist account of criminal punishment without giving up deontological constraints on using people as means to good ends and without invoking the desert of the criminal. He believes that the paradigm of using force in self-defense provides the key to a proper instrumentalist justification of punishment.

[1] Victor Tadros, The Ends of Harm: The Moral Foundations of Criminal Law (Oxford: Oxford University Press, 2011) (hereinafter Tadros).

Tadros's attempt to meet the challenge he sets for himself involves many steps. It is a huge understatement to point out that his is not an elegant theory of punishment. And like a watch mechanism with many cogs and wheels, Tadros's theory has many potential loci of problems that can produce a complete breakdown. And indeed, I believe his account falters fatally at several crucial points.

I. THE BARE-BONES THEORY

With some simplifications, Tadros's basic justification of punishment goes like this:

Step One: If D is about to harm V unjustifiably, V may harm D to prevent V's being harmed. (I leave aside the question of whether D must be culpable for this to be true, a question to which I shall later return. For now, I'll assume D *is* culpable.)

Step Two: If V cannot prevent D from unleashing the harm, V can use D as a means to prevent the harm. This is because D has a duty to prevent the harm he has unleashed, including a duty to use himself as a means to prevent it. In using D to prevent the harm, then, V is enforcing D's duty. (So, for example, if D has detonated dynamite in order to cause a boulder to roll towards V, V can shield himself from the boulder by throwing D in its path; for D has a duty to use himself as a means of averting the harm to V he has unleashed.)

Step Three: If D cannot prevent V from being harmed by what D has done, D's next best duty is to prevent V from being harmed by what someone else has unjustifiably done. (So, for example, if D_1 and D_2 have each started boulders rolling towards V, and D_1 can be used to stop D_2's boulder but not his, D_1's, boulder, D_1 has a duty to allow himself to be used to stop D_2's boulder.)

Step Four: If D cannot prevent V from being harmed by what D has done or prevent V from being harmed by what another has done, D's next best duty is to allow himself to be used to prevent others from being harmed unjustifiably.

Step Five: D's best method of fulfilling his duty at Step Four is to submit to punishment and thereby deter other potential Ds from harming other potential Vs unjustifiably.

That is how Tadros purports to provide an instrumentalist, deterrence-based justification of criminal punishment without endorsing thoroughgoing consequentialism. The fulcrum on which his theory rests is the duty of those who unjustifiably harm others to allow themselves to be used to prevent similar harms.

Of course, someone might object that the most obvious duty that an unjustified harm-causer has is to repair the harm he has caused – that is, to compensate his victim. Compensation, however, will often be inadequate to fulfill the duty. In some cases, the victim will be dead and thus beyond the reach of compensation. In many other cases, the harm causer will lack the resources to compensate his victim fully. In still other cases, the harm causer will resist compensating his victim. Punishment, therefore, will remain a necessity even in the presence of a duty to compensate.

And why is Tadros's theory not just an ordinary consequentialist one, one in which those punished are used as means to deter others and produce the maximum difference between the good consequences (less victimization) and the bad consequences (the suffering of those punished and others adversely affected thereby)? The reason is that Tadros believes in a deontological constraint on the pursuit of good consequences, namely, a constraint that prohibits using others as means to produce good consequences. And the reason deterrence-based punishment does not violate this constraint is because unjustified harms-causers – *but not others* – have a duty to use themselves (or allow themselves to be used) as means to prevent harms.

So, if Tadros's theory holds up, he has produced a non-retributivist justification of punishment without at the same time endorsing thoroughgoing consequentialism. He has found the elusive strait between the Scylla of retributivism and the Charybdis of consequentialism. But is there such a strait, and has Tadros really found it? My answers are 'doubtful' and 'no'.

II. DESERT-SKEPTICISM

Retributivism, of whatever stripe, depends on the existence of desert. The weakest variety just views desert as a ceiling on justified punishment but otherwise justifies punishment by its consequences. Stronger versions view giving people what they deserve as a positive good irrespective of its ancillary effects. All versions, however, depend upon desert's reality.

Tadros rejects retributivism largely because he rejects the reality of desert.[2] His rejection of desert is premised on the usual determinist-incompatibilist rejection of moral responsibility.[3] Although this is not the proper venue for arguing this point at length, I think that Tadros greatly underestimates the devastating implications of that position.

For one thing, I suspect that the position Tadros holds undermines the notion of wrongness and quite possibly normativity generally. I find it hard to imagine conveying the meaning of 'it's wrong' without implying the appropriateness of the reactive emotions of guilt, indignation, and blame as responses to wrongdoing. (That is why psychopaths fail to understand wrongness despite being able to identify wrongful acts.) And although this is more debatable, I suspect that the determinist-incompatibilist position threatens normativity of all types and substitutes for 'should do X' flat descriptive accounts of what will occur if X is done or not done and predictions about whether X will or won't be done.

Moreover, as I shall point out in several places, Tadros makes moves in his account that seem to make sense only if desert is real and is the explanation of the moves. And although Tadros claims to eschew reliance on desert, he does place a lot of emphasis on the significance of *choice*.[4] Yet one would think – although Tadros denies this – that the determinist-incompatibilist critique of desert would also undermine the significance of choice.

III. TADROS'S MEANS PRINCIPLE

As I said, Tadros endorses a deontological constraint against using people (without their consent) to produce good consequences.[5] (Tadros qualifies this constraint at both ends: he endorses a minimal Good Samaritan duty[6]; and he endorses threshold deontology,[7] which would allow the prospect of horrendous consequences to cancel the constraint.) The burden of much of his argument is to show that criminals have a duty arising out of their criminal acts that

[2] *Id.* at 60–87.

[3] *Id.* at 63–66.

[4] *Id.* at 52–58.

[5] *Id.* at 113–138.

[6] *Id.* at 129–138.

[7] *Id.* at 128.

cancels that deontological constraint and allows them to be used as means for the purpose of general deterrence.

I, too, endorse a means principle as a deontological constraint on the production of good consequences. For me, this principle prohibits the nonconsensual use of others' bodies, labor, and talents.[8] But my means principle differs from Tadros's in one significant way. My focus is on the causal path by which good consequences are produced. If an act's good consequences that otherwise would justify its bad consequences are brought about causally through the nonconsensual use of another's body, labor, or talent, my means principle is violated – *and it is violated irrespective of the intentions of the violator.*[9] On the other hand, if the good consequences are *not* the product of a causal path in which someone is used, my means principle is not violated – again, *irrespective of the intentions of those producing the good consequences.*

For me, then, intentions do not affect the permissibility of acts under the means principle. Indeed, intentions, as opposed to beliefs, do not even affect culpability. (They do affect our assessment of the actor's character, however).

Tadros believes intentions do matter to permissibility under the means principle.[10] Although I have elsewhere discussed at some length why I believe this is wrong, I shall here give one example that will illustrate my difference with Tadros and show why I believe his means principle is inferior to mine.

Let me take the now iconic hypothetical of Trolley. Tadros and I agree that one may permissibly switch the runaway trolley from the main track, where five lives are at risk, to the siding, where one life is at risk. (I also deny the relevance of whether one is deflecting a force already in motion or instigating a new force. Thus, I believe that if there is no siding on which to switch the trolley, but there is a barrier that can be lowered that will prevent the trolley from reaching the five but that when lowered will kill one worker trapped near the track, one may permissibly lower the barrier. I do not know Tadros's position on this point).

[8] *See*, e.g., LARRY ALEXANDER AND KIMBERLY KESSLER FERZAN, CRIME AND CULPABILITY: A THEORY OF CRIMINAL LAW (2009), 96–104.

[9] *See id.*; Larry Alexander & Kimberly Kessler Ferzan, "'Moore or Less' Causation and Responsibility", 6 CRIM. L. & PHIL. 81, 89–91 (2012).

[10] TADROS, *supra* note 1, at 139–166.

Tadros and I also agree that although switching the trolley is permissible to produce the good consequences – the saving of net four lives – because the one worker is not used as a means, pushing someone into the trolley's path to produce the same good consequences *is* a using of that person and is forbidden by the means principle.

Where Tadros and I part company is illustrated by the following variants of Trolley:

> Variant One: D is aware that switching the trolley will save a net four lives. But D doesn't care about that. His reason for switching the trolley is to kill the one worker on the siding, whom he hates.

Tadros would say that D violates the means principle in Variant One and thus that D's switching the trolley is an impermissible act. I disagree. The good consequences are produced by the switching irrespective of what happens to the one worker. He is not 'used' to produce those good consequences. A benevolently motivated person, knowing what D knows, could permissibly switch the trolley, which is why I deem D's act to be permissible. Indeed, because D is aware that his merely switching the trolley will save a net four lives, *D is not even culpable for switching it with the intent to kill the one.* He reveals an ugly character flaw by acting with that intention. But he does not act either impermissibly or culpably. He is in essence no different from an executioner who enjoys his work. His lawful executions are both permissible and nonculpable; and if we are to have executions, then we want there to be people who will do it, even if the only ones who volunteer have unsavory characters.

So I reject Tadros's concern with intention in construing the means principle. Indeed, *I reject the concern with intention, even if the actor's intention is to use another as a means*. To see why, consider a second variant of Trolley:

> Variant Two: Same as Variant One, except this time, D doesn't hate the one worker. Instead, D wants the trolley to kill the one worker so that the one worker's undamaged organs can be transplanted to two of D's friends, who will otherwise die of organ failure.

In Variant Two, although D knows that merely by switching the trolley he will save the five workers trapped on the main track, his reason for switching it is to use the one worker's body as a means for saving his two friends. Nonetheless, I believe D's switching the trolley is both permissible and nonculpable; Tadros, who would disagree with me about Variant One, would of course disagree with me about Variant Two.

Why do I think D's act is permissible and nonculpable? It is because a benignly motivated person, aware of what D is aware, would, like D, switch the trolley. He would do so to save the five workers. He would not do so to save D's two friends because, in the absence of the five threatened workers, his only reason to switch the trolley would be to use the one worker to save D's friends, an impermissible using. So the two additional lives D saves by switching the trolley do not count in the justification for switching because of the causal path by which those lives are saved. Nonetheless, because the switching is already adequately justified by its saving the five, a saving that occurs solely because of the switching and not because the one is killed – if he managed to escape, the five would still be saved, even if D's friends would not – D acts permissibly in switching no matter why he switches. And because he knows all the facts that make the switching permissible, he acts nonculpably in switching despite his intent to use the one. (He would be acting culpably, though permissibly, were he unaware of the five trapped workers; he would be attempting but not committing an impermissible act.)

Tadros believes that the causal path by which the good consequences are produced cannot be the basis of the means principle. In his Asteroid II example, an asteroid knocks a person off the footbridge over the trolley track, and his body stops the runaway trolley that would otherwise have killed the five workers.[11] He, however, is killed by the trolley. It would clearly have violated the means principle had we pushed the person onto the trolley track. Therefore, argues Tadros, we should feel worse about the death of that person than we feel about the death of the one worker in Asteroid I, where the asteroid knocks the trolley onto the siding, where it kills the one worker.[12] But we don't. We feel exactly the same way about the death in Asteroid II as we feel about the death in Asteroid I. Therefore, Tadros concludes, the causal path by which the five are saved cannot be what matters.

[11] *Id.* at 154.

[12] *Id.*

I believe Tadros is right that our reaction to the death in Asteroid II is the same as our reaction to it in Asteroid I. But that does not show that the causal path is irrelevant to the way *moral agents* produce good consequences. Deontological wrongs committed by moral agents can be consequentially beneficial. That is why consequentialists reject deontological constraints. We can cheer the asteroid on as it heads for the person on the footbridge even though it would be wrong for us to push him off. (And if we had a duty of easy rescue, and we could get the person off the track after the asteroid knocks him there but before the trolley reaches him, we would have a duty to do so, even though that would cost the five workers their lives. For our only reason for not fulfilling our duty of easy rescue and rescuing him would be his usefulness as a means, a reason upon which we must not act).

Nor can intention be the key to the means principle. For we can permissibly switch the trolley to the siding with the one worker *even if we intend to kill him as a means*. For switching the trolley produces justifying good consequences irrespective of whether it causes the one worker to be harmed.

So although I agree with Tadros that there exists a deontological constraint against using people as means, I disagree with his interpretation of that constraint. The causal path by which the good consequences are produced is what determines permissibility. The beliefs of the actor about that causal path and its consequences are what determines culpability. The actor's intention only reveals his character, contra Tadros.

IV. DEGREES OF CULPABILITY AND DEGREES OF HARM

An actor may be threatening a victim with great harm but be only minimally culpable for doing so. (I deal with completely innocent actors in Section V). Or an actor may be threatening a victim with trivial harm but be maximally culpable for doing so. (I deal with completely harmless but culpable acts in Section VI.) As an example of the first mismatch of culpability and threatened harm, suppose that D is about to dislodge, unintentionally but minimally recklessly, the boulder that will crush V. As an example of the second mismatch, suppose that a malicious D, intending to crush V to death, is about to dislodge a boulder that will in fact miss V almost entirely, except that it will bruise his big toe. How does Tadros's schema deal with the cases of low culpability/high harm and high culpability/low harm?

I do not think Tadros can handle these cases in any intuitively satisfying way. As I read him, in the low culpability/high harm cases, Tadros would say that V's using deadly force in self-defense – that is, preemptively in order to prevent D from dislodging the boulder – is perfectly justifiable.[13] After all, as we shall see, he believes deadly force can *justifiably* be used in self-defense against completely innocent aggressors.[14] So, of course, he would permit its use against minimally culpable ones.

Notice then that if deadly force can be used against the minimally culpable D at Step One, V should be able to use the minimally culpable D to shield himself from the boulder at Step Two. And so on and so on down the steps, leading to punishment of D quite out of proportion to D's culpability. Or so it would seem.

Now Tadros might say that is quite alright to inflict severe deterrent punishments on defendants who cause great harms but who are only minimally culpable for doing so. That will be a very hard bullet for Tadros to bite, however, as it would conflict with most people's intuitions about what is and is not just punishment.

Alternatively, Tadros might seek to distinguish the minimally culpable defendant who causes great harm from the maximally culpable one who does so. How could he do so? The usual route to take would be to say that the minimally culpable defendant does not *deserve* to be punished harshly no matter how much harm he causes. That route, however, is closed to Tadros because Tadros denies the reality of desert.

Another route that Tadros might take is to distinguish the culpability of risk taking from the culpability of purposely causing harm. Perhaps culpable risk takers cannot be used as means, unlike purposeful harm causers. This route, however, would foreclose a Tadros-like justification for punishing the reckless and knowing harm causers, another counter-intuitive result.

Tadros says very little about the extent to which the minimally culpable but maximally harmful defendants can be used and

[13] *Id.* at 175–181.

[14] *Id.* 246–256.

punished. What he does say suggests he believes such defendants can be punished to the extent they can be used to avert the harms they unleash – which means they can be punished severely despite their minimal culpability.

I think Tadros will have difficulty coming up with an intuitively plausible account of the justification for and limits of punishing minimally culpable but maximally harmful defendants. What about the punishment of maximally culpable but minimally harmful defendants? Presumably they can be used to prevent the harms they threaten. But can they be used in such a way that they will suffer a good deal more harm than they will cause?

Suppose, for example, that V's avoiding the boulder's bruising his toe by using D as a shield will result in great harm to D or even death. D, we said, did dislodge the boulder with the purpose of killing V, so it might seem entirely just for D to die even if only to save V's toe from a bruise. Indeed, I believe that only culpability, not results, should matter for punishment. So I would not object to harsh punishment for someone who tries to kill another but only ends up bruising his toe. But can Tadros endorse that conclusion in the absence of desert as a consideration? Does he endorse V's shooting D in defense of V's big toe? If not, then he cannot endorse punishment beyond the harm caused no matter how great D's culpability.

Again, Tadros does not say much on this point. What he does say, however, suggests that it is the amount of harm with which V is threatened that sets the limit on the amount of harm to which D can be manipulatively subjected, no matter D's level of culpability.

V. SELF-DEFENSE, PUNISHMENT, AND INNOCENT AGGRESSORS AND THREATS

In the last section I raised the question of how Tadros would deal with minimally culpable aggressors who threaten grave harm – harm disproportionately greater than their level of culpability – and who cause grave harm. Tadros is clear that the potential victim can justifiably use self-defense against minimally culpable aggressors who threaten grave harm. I agree with Tadros on this point. For example, I think that if Deborah is about to play involuntary Russian roulette on you just for the thrill of subjecting you to risks, you may use deadly force to stop her – even if her gun has many empty chambers and only one live round. The point generalizes to all culpable aggressors, no matter how low their level of culpability, if they threaten an act that *could* cause great harm. What one could justifiably do to the culpable aggressor to preempt his attack, however, does not for me determine the level of punishment that we are justified in inflicting if the attack is not preempted and occurs. For me, punishment is determined by the level of the defendant's culpability, which I believe is the measure of his retributive desert, and not by the harm he causes (or by the fact that he causes *no* harm).

Obviously, because he rejects retributive desert, Tadros cannot take my tack. One would think, as I suggested in the previous section, that he would endorse punishment measured by the harm the minimally culpable aggressor causes, even it if is much greater than his level of culpability. For would not the minimally culpable aggressor have a duty to prevent the harm he has unleashed (per Step Two), even if that would cost him his life? And would that not then justify severe punishment of the minimally culpable aggressor?

The limiting case of the minimally culpable aggressor is, of course, the innocent aggressor. Innocent aggressors come in a variety of forms – the young, the insane, the ignorant, and the out-of-control. I shall use one stock example as a stand-in for this variety, that of the innocent aggressor (IA) who is nonculpably ignorant of the harm he is about to cause. We can imagine that IA does not realize a villain has tampered with his cell phone so that when he places a call, a bomb will be detonated and kill V. Or we can imagine that IA, driving carefully, is nonculpably ignorant that there is a patch of black ice that will send his car out of control and kill V, a pedestrian. If V knows these facts, may he justifiably kill the IA in self-defense?

I have taken the position that although V may have an excuse or perhaps an agent-relative justification for killing IA in self-defense, V does not have a full-blooded justification for doing so.[15] That is because if D had full-blooded justification, a third party could stand in V's shoes and justifiably kill IA. Yet there is no agent-neutral reason to favor V's life over IA's and thus no agent-neutral justification for the third party to stand in V's shoes. Therefore, V does not have a full-blooded justification for killing IA in self-defense.

[15] *See*, e.g., ALEXANDER AND FERZAN, *supra* note 8, at 134–151.

Tadros disagrees. Following Jeff McMahan, he believes V can employ deadly force against IA because the latter is *responsible* for creating the risk even if not culpable for doing so.[16] As I have argued elsewhere, however, this overlooks the reciprocal, Coasean nature of the risks.[17] It is true that the IA cell phone user and IA driver are, by merely doing what they are doing, creating risks to Vs. But if Vs are carrying guns and are prepared to use them to prevent IAs from creating risks, *then Vs are acting in a way that creates risks to IAs*. The situation is perfectly symmetrical. Thus, there is no reason to favor innocent Vs over innocent IAs – who will, if they are killed by Vs, themselves be innocent Vs of innocent IAs (Vs).

Tadros not only believes that, despite their ignorance of the dangers they are creating or their youth or insanity, IAs have duties to avert the harms they are threatening and thus may justifiably be harmed by others to avert those harms and enforce that duty. He extends this to purely innocent threats – for example, those whose bodies have been hurled by tornadoes at potential victims.[18] He argues that innocent threats, were they able, would have duties to move their bodies to avert the harms. And because they would have such duties, we can harm them to avert the harms they otherwise will cause.

I disagree. Whether or not a victim may have an excuse or personal justification for harming an innocent threat, I see no reason to believe that there is some agent-neutral justification for having the innocent threat rather than the victim bear the harm. That the innocent threat would have a duty to avoid harming the victim if only he could do so does not do the trick.

Suppose Tadros is correct and I am wrong. Suppose, that is, that V may *justifiably* kill an IA (or an innocent threat) in self-defense. What does this entail for IAs at Step Two and beyond? If an IA has innocently set a boulder in motion towards V, may V use the IA as a shield? If so, does this mean that IAs may be punished for the harms they cause if this would somehow be instrumental in reducing harms to Vs?

Tadros appears to think not. For he argues that IAs can permissibly be harmed *eliminatively* – that is, to eliminate their threat – but

[16] Tadros, *supra* note 1, at 187–191, 241–245.

[17] *See* Alexander and Ferzan, *supra* note 8 at 149 n. 73.

[18] Tadros, *supra* note 1, at 248.

they cannot permissibly be harmed *manipulatively*.[19] That is, IAs may not be used as means to prevent harms.

And why is that? Tadros is anything but clear on this point. Indeed, it is not even clear that he is consistent on this point. He does say in a couple of places that the IA, though responsible for the threat he poses and thus liable to be harmed to eliminate that threat, cannot be harmed manipulatively. That suggests the IA, unlike the minimally culpable aggressor, has no duty to use himself as a means to avert the harm he has unleashed. But we are not told why the IA has no such duty, especially given Tadros's view that the IA can be *justifiably* harmed in self-defense. One salient reason for immunizing IAs from being used as means (manipulatively) is foreclosed to Tadros, however. He cannot argue that unlike culpable aggressors, IAs do not *deserve* to be used as means. For Tadros rejects the reality of desert.

Tadros even equivocates about whether IAs may be harmed eliminatively. For in his example of *Soldier Rescue*, he denies that a third party may kill a heroic rescuer who is, in attempting to rescue someone, nonetheless inadvertently posing a deadly threat to him.[20] If third parties may not kill IAs justifiably, why may victims?

With desert out of the picture, I see no way for Tadros to distinguish IAs from minimally culpable aggressors, or minimally culpable aggressors from maximally culpable aggressors, when it comes to using them as means and ultimately to punishing them for instrumental reasons. My guess is that when it comes to degrees of culpability and to IAs, Tadros is haunted by the obvious differences in negative desert. But in rejecting retributivism on the basis of denying its linchpin, negative desert, Tadros cannot invoke that desert to make the distinctions he would like to make. And he would like to make those distinctions because he cannot banish the ghost of negative desert.[21]

[19] *Id.* at 245–246.

[20] *Id.* at 233.

[21] Tadros, in his response, says that I have endorsed punishment beyond what is proportional to culpability. He is presumably referring to my doomsday machine article. Larry Alexander, 'The Doomsday Machine: Proportionality, Punishment and Prevention', 63 MONIST 65 (1980). But I was at pains to emphasize there that *protective* measures need not be proportional to the wrongs against which they were aimed, whereas *punishment* must be proportional to the wrong.

VI. HARMLESS CULPABLE AGGRESSORS: ATTEMPTS AND ENDANGERMENTS

There is a whole category of acts now uniformly regarded as punishable about which Tadros has next to nothing to say. That is the category of inchoate crimes that consists of attempts, solicitations, conspiracies, and culpable endangerments: culpable acts that cause no harm. In a book that is 360 pages long and devoted to justifying criminal punishment generally, Tadros devotes only two pages to the justification for punishing inchoate crimes.

There is a reason Tadros has virtually nothing to say about inchoate crimes. That is because their criminalization does not seem explicable on the basis of self-defense. If D attempts to roll a boulder towards V, intending to kill V, but V knows the boulder will miss him, V cannot kill or harm D in 'self-defense' – for V is not in danger and thus cannot be defending himself. Likewise, if D is driving quite recklessly, but V sees that D will not crash into him, V cannot kill or harm D in self-defense. And because there is no Step One of self-defense, there cannot be the subsequent steps that lead to D's punishment. (Tadros does flirt at one point with the idea that culpable *persons*, as well as aggressors, may be used as means; however, he ultimately rejects (as does Jeff McMahan) the permissibility of killing an attempter who is no longer a threat in order to avert a threat posed by another).[22]

Tadros suggests that perhaps we *are* harmed by attempters and endangerers because they divert security resources away from D's who *will* harm us.[23] Moreover, they harm us by disrespecting those whom they attack or endanger.[24] I find this half-hearted justification of inchoate crimes completely unconvincing. Many failed attempts and endangerments 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. As for disrespect, that cannot by itself justify criminal punishment. Moreover, most culpable endangerers have no particular victims in mind; indeed, there may not be anyone around who is in

[22] TADROS, *supra* note 1, at 187.

[23] *Id.* at 326.

[24] *Id.* at 327.

fact endangered by them, even if they believe otherwise (which is why they are culpable).

Finally, Tadros says nothing at all about solicitations, conspiracies, and complicity that have not yet led to their target crimes. Can their criminalization be explained on the self-defense model? If not, would Tadros abolish them as crimes?

I conclude that Tadros's self-defense justification for criminal punishment cannot account for the criminalization of inchoate crimes. Although I myself favor eliminating incomplete attempts from criminal codes and favor rethinking the basis for criminalizing solicitation, conspiracy, and complicity, I would not decriminalize the latter category.[25] And I surely would not decriminalize completed attempts and culpable endangerments, which I believe lie at the core of criminality.[26] Tadros's failure to provide a convincing rationale for these crimes is, I believe, particularly damning.

VII. SOME MISCELLANEOUS POINTS

1. Tadros says nothing about the burden of proof in criminal trials. Presumably, he does not believe potential victims must believe beyond a reasonable doubt that they are about to be attacked in order justifiably to employ force preemptively in self-defense – though he never discusses just what probability of attack as assessed by the potential victims is the threshold for justified preemptive defense. Whatever that probability is, should it be the same for a criminal conviction? One might think that if justified criminal punishment derives from justified self-defense, the required probability of guilt should be the same as that of attack. But Tadros does not tell us what either probability is, or why.

One basis for establishing the requisite probabilities is not open to Tadros. He cannot be concerned with *undeserved* punishment or *undeserved* preemptive attacks. For Tadros, desert is nonexistent, so that nothing is either deserved or undeserved.

2. Another somewhat curious position Tadros takes is with respect to compensation. Suppose the defendant could have been harmed to X extent to avert the threat he posed. And suppose now that defendant can fully compensate the victim for the harm caused

[25] *See* ALEXANDER AND FERZAN, *supra* note 8, at 223–225.

[26] *See id.* at 23–51.

but in doing so will only be harmed to Y extent, where Y is less than X. Tadros argues that defendant can be harmed to X extent even if it makes the victim better off than had defendant never acted.[27]

This position makes some sense in a system governed by desert. Defendant may deserve to be harmed to X extent (though his desert would be a function of his culpability, which may be more than or less than the harm he causes). And if defendant gets the harm he deserves – X – that may benefit the rest of us such that we are better off for his having acted culpably. However, it is hard to see why this would be so in a desert-free system premised on self-defense.

3. One final point. Tadros distinguishes among justifications 'relative to the facts' (FR), justifications 'relative to belief' (BR), and justifications 'relative to the evidence' (ER).[28] I understand FR and BR. ER, however, is different. Why should we care if a defendant's belief in justifying facts is unsupported by his evidence for that belief? The obvious answer is that when it does, we are inclined to call the belief a negligent belief. Does Tadros believe negligence is culpable and that negligent defendants can be punished as means? If so, I have yet another bone to pick with him.[29]

The distinction between BR and FR, without the need for ER, is adequate for explaining Tadros's *Miners*, where flooding one of two mineshafts will either save all ten trapped miners or kill them all, whereas doing nothing will result in the certain death of one miner.[30] For someone who believes the chances are even that the miners are in a given shaft and thus that flooding a shaft poses a 50% chance of ten deaths, the nonculpable course of action is to do nothing and let one miner die. The fact that were he to flood mineshaft A, no one would die, and thus that flooding A is FR justified, is immaterial. And so, I would argue, is the fact that there was evidence available from which one could conclude that flooding A was safe – if that evidence did not lead the actor to alter the belief that the chances were even that the miners were in A.

Of course, from the standpoint of someone who believes, correctly, that the miners are in B, flooding A, which will cause no deaths, is preferable to doing nothing. So the actor who does nothing

[27] Tadros, *supra* note 1, at 287–291.

[28] *Id.* at 217–220.

[29] *See*, e.g., Alexander and Ferzan, *supra* note 8, at 69–85.

[30] Tadros, *supra* note 1, at 222.

is not acting optimally from an FR perspective or from the observer's BR perspective. But evidence is only relevant as it affects beliefs.

My review of Tadros's book has to this point been highly critical. I wish to end, however, on a laudatory note. The book is full of interesting and provocative arguments and supporting examples, and I agree with a great many of Tadros's points. My disagreements are primarily directed at Tadros's conclusions regarding where his arguments lead.

Perhaps Tadros's most important points from my perspective are his various worries about how we retributivists should measure someone's desert, especially since we advocate working desert-based punishments into a legal system that is otherwise not desert-based in its allocations of benefits and burdens. I call that problem for retributivism the 'meshing' problem, and I commend Tadros for highlighting it.[31]

I also commend him for raising the problem of coercing people to support a costly system the justification for which is seeing that people receive their negative deserts.[32] Is that an impermissible 'using' that goes beyond any minimal affirmative duty to rescue? That's a question about which we retributivists need to think more than we have.

ACKNOWLEDGMENTS

Larry Alexander thanks Kimberly Ferzan for her helpful comments.

University of San Diego School of Law,
San Diego, CA, USA
E-mail: larrya@sandiego.edu

[31] *Id.* at 60–61, 70–73.

[32] *Id.* at 79–83. In other words, we need to ask if people can be rightfully coerced to labor in order to provide the means to see that retributive justice is achieved when they themselves are not the source of the injustice.