Law and Philosophy (2017) 36: 225–254
DOI 10.1007/s10982-017-9288-2

PATRICK TOMLIN

INNOCENCE LOST: A PROBLEM FOR PUNISHMENT AS DUTY

(Accepted 4 January 2017)

ABSTRACT. Constrained instrumentalist theories of punishment – those that seek to justify punishment by its good effects, but limit its scope – are an attractive alternative to pure retributivism or utilitarianism. One way in which we may be able to limit the scope of instrumental punishment is by justifying punishment through the concept of duty. This strategy is most clearly pursued in Victor Tadros' influential 'Duty View' of punishment. In this paper, I show that the Duty View as it stands cannot find any moral distinction between the permissible punishment of the guilty and the permissible punishment of the innocent in extreme circumstances, therefore undermining one the key pillars of its intuitive appeal. I canvass several ways to respond to this problem, arguing that a rights (or claims) forfeiture theory which employs the distinction between rights forfeiture and rights infringement (or claims forfeiture and infringement) is the best solution.

I. INTRODUCTION

Our traditional theories of punishment – those based on utilitarianism and retribution – have implications which are counter-intuitive for many. Retributivism (in its purest forms) requires us to see an offender's suffering as good, appropriate, or warranted in and of itself. Utilitarian theories avoid this entailment – punishment is, in Jeremy Bentham's words, 'an evil, though necessary to prevent greater evils'[1] – but they have a hard time explaining what is wrong with punishing the innocent. If punishment is at best a necessary evil, then provided punishing the innocent does enough good, there are no resources within utilitarianism to object to that, or to distinguish between the punishment of the innocent and the guilty.

[1] Jeremy Bentham, *Theory of Legislation,* Second Edition (London: Trübner & Co., 1871), p. 360.

For those who find both entailments problematic, one way to try to get around them is to combine the most attractive features of these two traditional theories – namely, utilitarianism's view of punishment as a regrettable necessity, and retributivism's link between guilt and punishment. This produces a group of theories which I call 'constrained instrumentalist' theories. These theories side with Bentham in seeing punishment as justified instrumentally, by its extrinsic good effects, whilst agreeing with retributivism that punishment is (ordinarily) only justified when applied to the guilty. In order to avoid retributivist conclusions about the value or appropriateness of suffering in and of itself, constrained instrumentalists must provide us with a concept other than desert which will explain why we may only (ordinarily) use guilty persons' punishment in order to produce the extrinsic benefits that punishment is thought to produce.

One variant of constrained instrumentalism limits instrumental punishment by justifying it through the concept of duty. Instrumentalism requires us to use offenders' suffering as a means to achieving good effects. Many philosophers believe that using others as a means in this way is ordinarily wrong. Duty-based views say that we can permissibly use people's suffering in this way when the suffering we cause through punishment is suffering that the person has an enforceable duty to experience. This is an attractive route for two reasons. First, punishment is hard to justify. It involves doing things to persons that it would ordinarily be impermissible to inflict upon them. If people have an enforceable duty to do, or to experience, something, that would explain why these things may be done to them. This is a general point about moral theory – we can justifiably harm people in the pursuit of some end when they have an enforceable duty to take on that level of harm in the name of that end. Think of cases where a culpable wrongdoer has created a threat – we may harm him to avert that threat if, as seems plausible, he has an enforceable duty to take on that level of harm in order to avert the threat. Second, a more particular point about punishment itself: It is an intuitive thought that offenders owe a 'debt' to their victims or to society. Although the language of 'debt' is often employed by retributivists, they owe an explanation of how punishment alone

pays a 'debt'.[2] Instrumentalists, however, can show how the extrinsic benefits of punishments that they demand as part of the justification of punishment can make the repayment of debt through punishment something real, rather than a mere metaphor. It seems plausible that offenders have debts and thus duties to their victims and, perhaps, to wider society, and punishment may be a way for them to discharge those duties, for example by reducing crime through deterrence. We can back up the idea that offenders have such duties with this thought: If criminals who avoid punishment, for example by escaping from prison (without harming anyone), do something wrong, that must be not only because the community has a *right* to punish them, but also because the offenders have a *duty* to submit to punishment.

In this essay, I will argue that despite its attractiveness, this variant of constrained instrumentalism has a serious problem. This problem comes to light not when we think about the central case of punishing offenders, but when we think about some ways in which punishments – or harms very like punishments – may be inflicted on the innocent. I will argue for the following claims. First, justifying punishment through duty will sometimes, in extreme circumstances, allow the punishment of the innocent. I take this to be a clarificatory and sympathetic claim. Second, and more importantly, justifying punishment through duty fails to differentiate between the permissible punishment of the guilty and some permissible punishment of the innocent, which is counter-intuitive. Third, another constraining concept – namely, rights forfeiture (or claims forfeiture) – can do a better job, when coupled with the notion of rights infringement (or claims infringement). Finally, I examine whether rights forfeiture and duty theories may be combined. That is, while I am sympathetic to using duty to justify punishment, I think that rights forfeiture (or claims forfeiture) must play a central role in the theory, if we are to properly distinguish between the punishment of the innocent and the punishment of the guilty. Endorsing this combination requires us to accept some controversial positions on the relationship between rights (or claims) and duties.

[2] As Warren Quinn observes: 'In punishment . . . there is a "taking away" from the criminal without any obvious transfer of what is taken away to anyone else.' Warren Quinn, 'The Right to Threaten and the Right to Punish' in *Philosophy & Public Affairs* 14 (1985): 327–373, at p. 334, n. 11.

What I have to say applies to any theory of punishment which is both instrumentalist and uses the concept of duty to justify instrumental punishment.[3] However, in this essay I will focus on the theory which most clearly and directly appeals to duty as its central justifying concept – Victor Tadros' 'Duty View'.[4] What it has to say about the innocent is important ground on which to confront this theory: Tadros is explicit that his theory's protection for the innocent is its main advantage over non-constrained instrumentalist theories of punishment.[5] As such, the problems I expose in the theory are problems by the theory's own guiding lights.

II. PUNISHING THE INNOCENT: CONCEPTUAL AND NORMATIVE ISSUES

Most of us have the conviction that punishing the innocent is, to use a vague phrase, morally troubling. However troubling we find the idea of punishment generally (and we should find it at least prima facie troubling[6] – even if it can be justified, it involves the state intentionally harming its own citizens), the idea that someone who hasn't broken any law, nor done anything morally wrong, ends up being punished (or having an experience identical to punishment) is even more troubling, and even harder to justify. Even if and when

[3] An important theory that is related is the self-defence-based view, proposed by Daniel M. Farrell and Phillip Montague. These views do not appeal first and foremost to duty, but rather to the *right* of Victim to punish/defend herself in situations in which either Victim or Aggressor must be harmed. However, should Victim choose to harm Aggressor, Aggressor surely has a duty to accept that harm (and would not, for example, have any right of self-defence against the harm). Therefore, self-defence views which appeal to duty *and* which recognize duties of rescue potentially have similar problems. They can (as I suggest later) appeal to rights forfeiture and rights infringement to differentiate the cases. Farrell points to rights forfeiture as an important concept. See Daniel M. Farrell, 'The Justification of General Deterrence' in *Philosophical Review* 94 (1985): 367–394; Phillip Montague, 'Punishment and Societal Defense' in *Criminal Justice Ethics* 2 (1983): 30–36.

[4] The central statement of this theory is in Victor Tadros, *The Ends of Harm: the moral foundations of the criminal law* (Oxford: Oxford University Press, 2011). Hereafter, all parenthetical page references are to *The Ends of Harm*. Tadros' theory has attracted a great deal of scholarly attention, but none of the respondents has thus far noticed the problems I expose here. See the symposia (and Tadros' responses) in *Law and Philosophy* 32(1–3) (2013), *Criminal Law and Philosophy* 9(1) (2015), and *Jerusalem Review of Legal Studies* 5 (2012).

[5] On p. 42, Tadros states that: 'The aim [of the book] is develop an instrumentalist account of punishment that can meet the objections that retributivists have raised to consequentialist theories . . . It shares with retributivism the idea that respect for individuals, rather than the fact that it would be futile or counterproductive, requires us not to punish the innocent or to punish the guilty disproportionately.' In fact, as we shall see, Tadros denies that retributivism itself *necessarily* respects this shared idea.

[6] Sometimes 'prima facie' is used as a synonym for 'pro tanto'. I do not mean it in that sense here. I mean to say that we have cause for thinking punishment troubling, on the face of it. It may be that (justified) punishment is *in no way troubling*, which would not be the case if it were pro tanto troubling (for then its troublingness might be outweighed, but would not be extinguished).

punishing the innocent is permissible, there is a fundamental moral distinction between punishing the innocent and punishing the (legally and/or morally) guilty.

It is worth emphasizing four different ways in which an innocent person may be 'punished'. The first is the way that is most familiar to us in the real world: We try very hard to establish the facts of the case, but come to a mistaken conclusion. Call this accidental punishment. Any actual system of criminal law, no matter how well designed and implemented, will have some accidental punishment.

The second is the kind that will be familiar from objections to consequentialist theories: where some person or group of persons subject someone they know to be innocent to punishment, or an experience that is, in all relevant respects, identical to punishment, because doing so will achieve some good effects. Some claim that it is a conceptual impossibility to *punish* innocent persons in this way.[7] It is claimed that this cannot be punishment since those who do the 'punishing' do not intend the suffering of the person to be an honest punitive response to past wrongdoing. H.L.A. Hart's famous definition of punishment requires it to be distributed to 'an actual or supposed offender *for his offence*',[8] and this form of imposed harm does not fit this definition. Hart claims we should still call this punishment – albeit punishment of a 'secondary' kind. Rawls calls this 'telishment'.[9] I will not try to settle these conceptual issues here, but I will use the term telishment to refer to this particular kind of 'punishment' of the innocent. Like Tadros,[10] as far as I am concerned, however we settle the conceptual issue, the normative issues remain the same: Under what conditions (if any) can telishment be justified? And how should we think of justified telishment compared with justified punishment? Tadros takes his theory to show that telishment is unjustified. I will show that in fact it justifies it in some circumstances. And I will show that, like consequentialist theories,

[7] I am grateful to Doug Husak and Alec Walen for encouraging me to address this.

[8] H.L.A. Hart, *Punishment and Responsibility*, Second Edition (Oxford: Oxford University Press, 2008), p. 5. My emphasis.

[9] John Rawls, 'Two Concepts of Rules' in his *Collected Papers*, Samuel Freeman ed. (Cambridge, Mass: Harvard University Press, 1999), p. 27.

[10] On p. 313 n.1, Tadros references the concept of telishment. Tadros is indifferent as to whether we call this punishment of the innocent or telishment, since, for him, the normative issues remain the same, whichever we call it.

Tadros' theory, as it is presented, cannot distinguish between justified telishment and justified punishment.

The third way to punish an innocent person is what we can call framing: where some person or group of persons knows the defendant to be innocent, but makes it look like they are not, such that those who end up convicting and punishing them believe the defendant to be guilty of wrongdoing. Therefore, those punishing act and think in exactly the ways that they do under accidental punishment, and the experience is identical for the person on the receiving end. Some may want to deny that this is really punishment as well, and call it a form of telishment. But it is hard to see why. Since those who punish do so in response to perceived wrongdoing, they punish a *supposed* offender *for* his supposed offence. While some others (including the defendant) know of his innocence, that will often be the case in accidental punishment as well, and nobody wants to deny that *that* is punishment.

The final form of punishing the innocent comes about when we are indifferent, or insufficiently attentive, as to whether someone is guilty or innocent. Imagine we *suspect* someone to have committed an offence, and so seek to punish them for that offence. This would be punishment (we punish a supposed offender for his offence), and so, if the offender turns out to be innocent, it would be a form of accidental punishment. But our indifference toward their guilt would be most concerning. Let's call this indifference punishment. Saul Smilansky has persuasively argued that whilst telishment is the form of punishing the innocent that is usually used to object to consequentialist theories, indifference punishments are just as much of a concern.[11]

For the purposes of this essay, we are interested in telishment, framing, and indifference punishments (i.e., not accidental punishments). For ease, I will refer to these, collectively, as 'the punishment of the innocent', notwithstanding the conceptual controversies I have already alluded to. I take it that we are, to put it mildly, morally concerned about all of these. Let's call this concern the *Innocence Intuition*. I will state it here as broadly as I can, because, like many of our moral intuitions and convictions, it is merely a starting point for philosophical inquiry and, in my own pre-theoretical judgments at

[11] Saul Smilansky, 'Utilitarianism and the Punishment of the Innocent: the general problem' in *Analysis* 50 (1990): 256–261.

least, is held in a vague way. I have also phrased it as an evaluative intuition. This will be important as we go through, but it is, I think, a natural way to phrase the initial intuition. It is appropriate to confront Tadros in particular on this evaluative ground, since his principal disagreement with retributivism concerns not what it says about what we should *do,* but rather what it says about how we should *evaluate* the punishment of the guilty.[12]

The Innocence Intuition There is a fundamental moral distinction between punishing the guilty, on one hand, and the harms done to innocent parties under telishment, framing, and indifference punishments, on the other. All else equal, it is always morally worse to punish the innocent than to punish the guilty.

We can draw three different kinds of lines when we think about punishment. The first are factual, or descriptive, lines. Here we will focus on one such line – the one between guilt and innocence.[13] The second are evaluative lines which morally distinguish different kinds of punishment and what we should say and think about those punishments (for example, whether incidences of punishment are good or regrettable). The third are normative lines, which we draw between permissible and impermissible punishments. The Innocence Intuition draws an evaluative line, which tracks our factual distinction between guilt and innocence. It notes two differences between the two types of punishment: one of type and one of degree. In terms of degree, all else equal, punishing the innocent is always worse. In terms of type, it says (vaguely) that the distinction between punishing the innocent and punishing the guilty is *fundamental*. Punishing the innocent is not merely worse than punishing the guilty – it is a distinctive form of badness. We will hope that our theories of punishment, which primarily tell us when and why punishment is permissible, will heed the Innocence Intuition, both in that they will offer robust protection against punishment for the innocent, and that they will explain and validate the initial evaluative intuition.

[12] Tadros writes: 'I reject retributivism primarily because after reflecting on the lives of wrongdoers I simply lack any conviction that their suffering is good.' ('Responses' in *Law and Philosophy* 32 (2013): 241–325, at p. 259).

[13] In order to set aside the issue of whether it is moral or legal guilt/innocence that matters, for the purposes of this paper 'the guilty' have committed an act which is morally wrong, is morally permissible to punish *and* has been criminalized. If we focus on moral guilt, this 'factual' line will have normative commitments built into it.

While some accidental punishment is inevitable in any system of punishment, some may say that the kinds of punishment of the innocent we are interested in here – telishment, framing, and indifference punishment – are never permissible. They might endorse something like this:

The Strong View It is always impermissible to telish or frame someone, or to punish someone we merely suspect of having committed an offence.

Of course, telishment, framing, and indifference punishments are clearly something that any actual system of punishment must be designed to avoid. But there are extreme cases in which many of us would endorse them. Imagine, for example, that a group of Ruth-haters contacts the government and demands that Ruth be given one month in prison. If she is not, they will carry out a terrorist atrocity, with huge loss of life. Ruth is innocent. The Ruth-haters are only demanding a very small amount of punishment. Since it is known that Ruth is innocent, this would be telishment. Telishment is very hard to justify, but if the stakes get high enough, as in this case, can't it be justified in extremis? We can modify this case to make it a framing case. Imagine that the terrorists believe Ruth to have committed an offence, and are demanding she is mildly punished for this offence. Government officials can make it look as if Ruth did indeed commit the offence, so as to trick police, jury, court, and public into believing she committed it. They will then punish her. Again, this kind of conduct looks like it may be permissible if the stakes are high enough, and the punishment low enough. Finally, the case can be amended to make a case of indifference punishment. Imagine the government suspects Ruth of a crime, but cannot prove it beyond reasonable doubt. The terrorists believe she committed the crime, and are demanding she be punished for it. We might be prepared to punish her for the crime, even though we are unsure of her guilt or innocence.

If we endorse punishment in these cases, but want to continue to pay attention to the Innocence Intuition, we might endorse this (rather vague) normative principle:

The Moderate View Telishment, framing, and indifference punishments are only permissible in extreme circumstances. The good produced by these punishments must vastly outweigh the harm done to the punished person.

Even though, in extreme circumstances, it allows us to punish those we know to be innocent, a theory which leads us to the Moderate View can nevertheless heed the Innocence Intuition in two senses. Firstly, the Moderate View requires circumstances to be exceptional before punishment of the innocent is permitted – it does not take punishment of the innocent lightly, and sees it as harder to justify than punishment of the guilty. Secondly, even when punishment of the innocent is allowed, it allows us to retain (i.e., does not directly contradict) the *evaluative* thought that there is something that is troubling or regrettable about this punishment that is *not* there when we punish the guilty. The permissible punishment of the innocent can remain a different moral kettle of fish from the permissible punishment of the guilty.

Traditionally, the punishment of the innocent is a stick with which to beat instrumentalists – those who seek to justify punishment by its extrinsic good effects. For example, consequentialists who believe that punishment is justified whenever it does more good than harm, must believe that whenever punishing the innocent does enough good to outweigh the harm to the punished, it is justified. Moreover, provided the good effects are the same, punishing an innocent person would be morally identical to punishing a guilty person. Therefore, certain kinds of traditional consequentialist theories (such as utilitarianism) do not appear to vindicate or explain the Innocence Intuition, nor can they necessarily justify the limits we want to see our criminal justice institutions respect. To the extent that they *can* justify such limits, they appear to do so for the wrong reasons – because punishing the innocent is ineffective, not because there is something fundamentally morally troubling about it.

This has led many theorists to believe that the Innocence Intuition, and the related normative positions explored above, speak in favour of retributivist theories of punishment. It is clear that retributivism can vindicate and explain the Innocence Intuition. Because retributivists believe that the guilty deserve punishment, while the innocent do not, right at the heart of the theory is the fundamental moral distinction between the two types of punishment that the Innocence Intuition requires. Even if a retributivist concludes that innocent Ruth can be punished in order to avert a terrorist atrocity, she can nevertheless hold on to the idea that there is a

fundamental distinction between this punishment (undeserved, and therefore bad, but permissible) and punishment of the guilty (deserved, and therefore, at least pro tanto, good). That is, even when she is prepared to shift the *normative* line that divides permissible and impermissible punishment away from the *factual* guilt/innocence divide, an *evaluative* line may remain on the guilt/innocence divide.

Yet precisely because these evaluative and normative lines can come apart, the retributivist is not *automatically* in a better position than the deterrence theorist when it comes to how much punishment of the innocent they would allow. As Tadros clearly shows, what he takes to be the basic retributivist claim – that punishment is intrinsically good when and because it is deserved suffering, and is intrinsically bad when and because it is undeserved suffering[14] (i.e., the evaluative line corresponds to the factual one) – does not, on its own, lead us to either of the normative view articulated above, and thus will not *necessarily* place tight limits on the permissibility of the punishment of the innocent (pp. *35–37*).

So, retributivists are on firm ground with the Innocence Intuition, but must show that their theories lead us to plausible normative limits on the punishment on the innocent (which is not, of course, to say that they cannot or have not done so). In what follows I will argue that the Duty View finds itself in the opposite position. It can place sensible normative limits on the punishment on the innocent but, in doing so, it cannot, on its own, endorse or explain the Innocence Intuition. This is, however, somewhat obscured from view in the presentation of the Duty View, since Tadros tends to write as if he endorses stricter limits on the punishment of the innocent (namely, the Strong View) than he actually does, which allows him to draw a bright *normative* line between innocence and guilt. This bright normative line carries a kind of evaluative line with it – permissible punishment and impermissible punishment should, of course, be evaluated differently.[15] However, once we shift the normative line to be in accordance with Tadros' *actual* views, we

[14] Tadros' characterisation of retributivism is controversial. In particular, it does not incorporate 'negative' or 'limiting' retributivism. Larry Alexander claims that it is 'easy as pie' to justify punishment on grounds *like* Tadros' if one makes use of desert as a limitation on punishment. See his 'Can Self-Defense Justify Punishment?' in *Law and Philosophy* 32 (2013): 159–175, at p. 159. One way to read Tadros' project, and related views, is as an attempt to develop a position that will provide similar conclusions to the negative retributivist position, but without relying on the concept of desert.

[15] Larry S. Temkin, *Rethinking the Good* (Oxford: Oxford University Press, 2012), p. 10, n.8.

find that (unlike the retributivist) his evaluative line shifts with it, leaving Tadros' theory with no normative *or* evaluative line between the innocent and the guilty.

III. THE DUTY VIEW

Tadros' argument for the Duty View begins with a rejection of the retributivist claim that the suffering of offenders is intrinsically good. Like all instrumentalists about punishment, Tadros sees punishment as a necessary evil. But Tadros aims to present an instrumental theory which provides the tight limits on punishment that consequentialists fail to provide. Tadros declares that a major advantage of his theory is its position on the punishment of the innocent, which he repeatedly claims his theory declares to be wrong and thus impermissible.[16] Thus, Tadros writes as if the Duty View endorses the Strong View on the punishment of the innocent.

The theory which will deliver these limits stems from an analysis of the 'means principle'. Instrumentalist theories of punishment, and especially theories like Tadros' which appeal to general deterrence, are often objected to on the grounds that they *use people as a means* to some end.[17] Tadros endorses the means principle (the principle which states that it is usually wrong to use people in this way), but argues that it has important limits. Forcibly using people is not *always* wrong: 'It is often permissible to harm a person as a means to a greater good, I claim, if that person would themselves have a duty to pursue that good even if doing so would harm them to that

[16] Here are a series of claims that Tadros makes about punishing the innocent: the innocent *may not* be punished (p. 40); punishing the innocent is (or seems) *unjust* (pp. 42, 314); the Duty View '*requires* us not to punish the innocent' (p. 42); 'Most people think that…punishment of the innocent…*violate[s] basic moral requirements* – doing these things is *fundamentally unjust* rather than merely imprudent' (p. 138); the Duty View shows how it is permissible to 'punish *the guilty* for reasons of general deterrence' (p. 138); punishing the innocent 'seems *wrong*, even if this will prevent greater harms, because it would harm the innocent person as a means to the greater good' (p. 265); 'The idea [of the Duty View] is to show that although it is generally wrong to harm a person as a means to the good of others, *explaining why punishment of the innocent is wrong*, it may be permissible to harm *offenders* as a means to the good of others' (pp. 265–266) [furthermore, for Tadros, if some conduct is wrong, that is 'a *morally decisive* reason not to do it' (p. 217)]; 'committing a crime is a *necessary condition* of punishment' (p. 276) (all emphases mine). In 'Responses' Tadros continues to write as if his theory only justifies the punishment of guilty – in his summary of the Duty View (p. 242) he implies that only 'offenders' are liable to punishment. In his more recent *Wrongs and Crimes* (Oxford: Oxford University Press, 2016) Tadros appears more relaxed about the punishment of the innocent, though, importantly, he has in mind criminalizing innocent conduct, rather than punishing those innocent of criminal conduct. See, especially, chs. 6 and 17.

[17] Daniel M. Farrell also accepts this as the major challenge to instrumental theories. See Daniel M. Farrell, 'The Justification of Deterrent Violence' in *Ethics* 100 (1990): 301–317, at p. 301.

degree' (p. 117). Tadros adds that the duty must be enforceable, and whilst not all duties are enforceable, when a duty is grounded in the avoidance of very serious harm to others, it will typically be enforceable (i.e., we will have no right to do wrong) (p. 131). Tadros' thought is that what is wrong about harming a person as a means to an end is that we use them in the service of some goal. This is permissible when they are morally *required* to have taken that as their goal at the same cost (p. 129).

Tadros argues that punishment can be justified as an enforceable duty to deter future wrongdoing. Offenders, says Tadros, have duties to their victims. They have failed in their primary duty – not to harm their victim – and so acquire secondary duties, duties to do the next best thing. And these duties require that they protect their past victims from future harms, if they can do so at proportionate cost (pp. 265–277).[18] One way that they can protect their past victims from future harms is through being punished, if this means that others are deterred from harming their victims in the future. Tadros goes on to argue that, where they cannot protect their own victim specifically, offenders have (enforceable) duties to form arrangements that enable them to protect one another's victims, and that rights to be protected can be transferred from victims to their friends, family and fellow citizens (pp. 280–281). If both these extensions to offenders' liabilities (from specific victim to victims generally, and from victims generally to citizens generally) are valid, then we get from an initial duty of protection owed by the offender to his victim, to a duty to protect society generally. Where punishment protects society, it is permissible if the costs of punishment do not exceed those that the offender had a duty to accept.

One important element of his theory that Tadros finds attractive is this: The harms that punishment imposes are bad and stay bad. Unlike retributivism, where the harms of punishment can be transformed into a good through desert (which Hart called the 'moral alchemy' of retributivism[19]), on the Duty View people may become liable to harm through their choices to engage in criminal conduct, but the harm

[18] At this point, Tadros' theory differs in important ways from Daniel Farrell's related self-defence based account, as for Farrell, we can only punish in order to avert threats against victims which offenders themselves have created, whereas for Tadros the duty is a duty to protect victims against any threats they face. See: Farrell, 'The Justification of General Deterrence'.

[19] H.L.A. Hart, *Punishment and Responsibility* (Oxford: Oxford University Press, 1968), pp. 234–235.

remains a necessary evil. As Tadros puts it, we can recognise that offenders have duties to be harmed which others do not, 'But in recognizing that they have such a duty we need not appeal to the idea that harming them will be in any way good, or else less bad, than the harm that would be inflicted on other people' (p. 53).

IV. PUNISHING THE INNOCENT: THE NORMATIVE STANCE

In the previous section, I showed that Tadros often claims that his theory supports the Strong View – the view that punishment of the innocent is always impermissible. In this section I want to show two things. The first is that Tadros' theory, at its most general level, doesn't lead us to any particular conclusion about whether or when punishing the innocent is permissible or impermissible. The second is that if we look to Tadros' statements about various other things (including, in particular, duties of rescue) his more specific version of the Duty View *doesn't* rule out punishing the innocent. Rather, it will most likely support a (better specified) version of the Moderate View, which allows for the punishment of the innocent in exceptional cases.

This second claim is, I believe, both clarificatory and sympathetic. It is developed from Tadros' own arguments and I think that I show that Tadros' specific version of the Duty View actually delivers *more* plausible restrictions on the punishment of the innocent than those he often claims for it. However, since Tadros places such weight on his theory delivering the right restrictions on punishing the innocent (this is, after all, what distinguishes it from a pure consequentialist deterrence theory), and often declares that punishing the innocent is impermissible, it is important to show both that his *general* theory doesn't have a concrete position on if and when we can permissibly punish the innocent (indeed, it is fully compatible with the consequentialist theories that he aims to provide an alternative to), and that his specific version of the theory actually *allows* the punishment of the innocent.[20]

[20] Others have taken Tadros at his word on this. In his paper on Tadros' theory, Douglas Husak clearly takes Tadros to have argued that only the guilty are liable to punishment. He writes: '[H]ard questions about the moral status of punishing the innocent should be posed to all theorists . . . Retributivists can ask Tadros: why confine punishment to those said to be *liable* to it?' (Douglas Husak, 'Retributivism *in Extremis*' in *Law and Philosophy* 32 (2013): 3–31 at p. 16, emphasis in original). I think Husak's presumption (that Tadros means to confine punishment to the liable, and that only the guilty are liable) is an understandable way to read Tadros (see note 16). I seek here to show that it is nevertheless the wrong way to read him.

To see how limits to the use of punishment cannot be read straight from the general version of Tadros' view, here is a reconstruction of the general idea of the Duty View.[21]

DV1 Harming someone as a means to the greater good is usually wrong.

DV2 An exception to the means principle (*DV1*) occurs when the person has an enforceable duty to take on a given level of cost (harm) in order to achieve some end. It is then permissible to impose those costs (harm them to that degree) if doing so is necessary to achieving that end.

DV3 Therefore, punishment is permissible harm when and because a person has an enforceable duty to assist people up to a certain cost.

DV4 Offenders acquire such duties through past wrongdoing.

What does this tell us about the limits to punishment? Nothing. Tadros focuses our attention on the duties that criminals have toward their victims (*DV4*), but this tells us nothing about the limits to punishment because it tells us about one group of people who have duties to incur costs (be harmed), and says nothing about who else has such duties, or does not have such duties. It would be possible for me to sign up to the Duty View in this general form, and have just about any position on the punishment of the innocent: that telishment is never justified, is justified in highly constrained circumstances, or is justified whenever it will do more good than harm. It all depends on what enforceable duties the innocent have, and the view as it is presented is silent on that (other than general assertions about the Duty View not allowing the punishment the innocent (see note 16)). So, just as the general view of retributivism cannot necessarily protect the innocent (pp. 35–37), neither can the general Duty View.

That the Duty View in its general form can support punishing the innocent whenever it will do more good than harm is worthy of note. Tadros claims that his theory is non-consequentialist and this claim appears to be based on his endorsement of, and his theory preceding from, the non-consequentialist means principle. But Tadros amends the means principle in such a way that it becomes completely compatible with consequentialism. Consequentialists (at least as traditionally understood) cannot endorse the principle that we must *never* treat a person as a means, for they will think it permissible to (for example) push someone off a bridge in order to

[21] See also Tadros' own summary in 'Responses', p. 242.

stop a trolley that will otherwise kill five people. But the consequentialist *can* endorse the principle that we may only treat a person as a means when they have (or would have) a duty to act in that way, because (for example) they will believe that the person on the bridge has a duty to jump, or to allow themselves to be pushed (since it will do more good than harm). Therefore, Tadros' 'non-consequentialist' principle is amended or clarified in such a way as to make it perfectly compatible with consequentialism. This is not to say that Tadros' view is a consequentialist one, but that the non-consequentialist character of the theory does not stem from the amended means principle. Instead, the substantive claims about who has what duties to be harmed provide the non-consequentialist character of the theory.

Moving on to Tadros' more specific view on duties, Tadros' official view is that the Duty View endorses *The Strong View* on the punishment of the innocent. Yet Tadros often reminds us that we sometimes have duties to bear costs in order to avert harm to others *even when we are not responsible for the creation of the threat and have not harmed anyone else*. For example, Tadros argues that if I could save your life at minimal cost (for example, sustaining some bruising), then I have an enforceable duty to sustain the harm (p. 129). This means that others can force me to sustain these costs in order to achieve the same end. Given these enforceable duties of rescue, surely Ruth has an enforceable duty to be mildly harmed (telished) in order to avert the terrorist threat. Similarly, since Ruth would have an enforceable duty to confess to a crime she didn't commit (in order to avert disastrous consequences), framing her would also appear to be permissible. Finally, consider the case in which we suspect Ruth of having committed the crime, but in which the Ruth-haters are demanding punishment. In such a case, Ruth has an enforceable duty either way – either she committed the crime and has a duty in accordance with *DV4*, *or* she didn't and has a duty in accordance with the duties of rescue Tadros affirms. Therefore, we can harm Ruth through punishment even though we only suspect her of guilt, since we can be sure that she has a duty to be suffer the harm.

Therefore, to the initial description of the Duty View, we can add the following claim, which makes it compatible with the Moderate View:

DV5: When the goods vastly outweigh the costs, the innocent have enforceable duties to be harmed in order to assist others. This means that telishment, framing, and indifference punishment can, in rare circumstances, be justified.

As I have made clear from the outset, when the punishments are mild enough and the stakes high enough, it seems plausible that the innocent can be permissibly punished. And it seems plausible to say that, in such circumstances, the innocent have a duty to submit to punitive institutions. For example, Ruth would act wrongly if she absconded and allowed the terrorist atrocity to occur. So I think the fact that a specified version of Tadros' theory can lead us to this conclusion is a welcome result for that theory – this claim is sympathetic and clarificatory. However, even if Tadros were to be upfront about this, and modify his stance by saying that the punishment of the innocent is *ordinarily* impermissible, we find a new problem: Once we are clear that the Duty View's normative line does not lie exactly along the guilt/innocence divide, this has serious ramifications for the theory's evaluative claims.

V. PUNISHING THE INNOCENT: THE EVALUATIVE STANCE

The real worry about the Duty View (and duty-based theories of punishment in general) is how they *evaluate* the permissible punishment of the innocent. Consider, again, the Ruth case, and compare it with this case: Sarah is required to endure punishment, thus deterring others, because she murdered someone. She is given fifteen years in prison, which prevents three murders. Recall that, on Tadros' view we are not to see the suffering of wrongdoers as *good*, but rather as a *necessary evil* justified by the good effects that it has, and the fact that the person has a duty to bring about those good effects. The suffering of the guilty is not good, nor less bad, than the suffering of the innocent (p. 53).

Imagine that you are a guard walking down the hallway between Sarah and Ruth's cells, and you are one of the small number who has been made aware of Ruth's innocence. Of course, you should feel differently toward Sarah than you do toward Ruth – Ruth is innocent, whilst Sarah has committed a horrendous wrong. But according to the Duty View you should view their *suffering* in the same way, as morally equivalent. Both are suffering, which is bad in both cases. But both are suffering because it's their duty, so their

punishment is permissible. On Tadros' theory as it stands, there is, so far as I can see, nothing more to say. And this is true of any view which couples Bentham's view of punishment as a necessary evil with the idea that duty can justify that evil being imparted on the innocent. The two punishments are morally equivalent, there is no significant moral distinction between them.

The Innocence Intuition, as I initially formulated it, made two claims. First, there is a fundamental moral distinction between punishing the innocent and punishing the guilty. Second, when all else is equal, punishing the innocent is worse than punishing the guilty. Ruth is suffering mild harm in order to avert a very serious threat, whilst Sarah is suffering a major harm to avoid a lesser threat, so the cost-benefit ratios of the two punishments are importantly different, and so all else is not equal. Since all else is not equal, the Duty View does not (as yet) conflict with the second claim, but it does conflict with the first part of the Innocence Intuition. For retributivists, there is always a fundamental moral difference between punishing the innocent and the guilty, even when both are permissible. For Tadros both are evaluated in the same way – the harm is bad but the person has a duty to suffer it.

A defender of the Duty View may claim that she has no trouble seeing these two punishments as morally equivalent – Ruth's minor suffering is doing so much more good than Sarah's major suffering. But what if we alter the variables? Let Ruth's punishment remain as it was: a minor punishment to avert a very serious threat. But let's also say that Sarah and Ruth live in an exceedingly law-abiding society, and that there is only one other threat, and it is of an equally serious nature. However, in order to avert it, we need only punish Sarah a little bit. The exact same amount as Ruth, in fact. Since, according to the Duty View, permissible punishment is a duty to suffer *necessary* harm, then we cannot punish Sarah any more than the tiny amount required to avert the threat. Now the cost-benefit ratios between the two cases are uniform. But the Duty View still views them as equivalents. Being guilty may raise the amount of harm you potentially have a duty to incur to in order to avoid harm (i.e., we can do more to Sarah to avoid the same harm than we can to Ruth), but *within the boundaries of harm to which we potentially have a duty to bear* all harm is, for Tadros, on a par – permissible but

regrettable. Therefore, now with all else equal, the Duty View *still* cannot differentiate between the cases, and so the Duty View clashes with the second part of the Innocence Intuition.

A defender of the Duty View might be tempted to claim that a significant difference is to be found in the fact that Ruth's punishment is nearer the potential maximum harm she could be permissibly forced to bear, whilst Sarah is only suffering a tiny proportion of the harm to which she is potentially liable. But it is hard to see why this would lead us to evaluate the harms suffered by the two persons differently from one another. That would require us believe that harms are more troubling the nearer they are to the person's maximum potential duty. But that would introduce distinctions in harming *within* permissible harms, which would seem to conflict with Tadros' basic idea that all permissible harms are just as bad as each other (p. 53).

When Tadros rejects retributivism, he focuses on the way in which it transmutes harm from ordinarily bad to sometimes good. This causes Tadros to reject the 'moral valence' of harm – the idea that our evaluations of harm can shift according to the moral situation of the person who suffers them. But our rejection of the moral valence of harm can take two forms – a stronger and a weaker form. These are as follows:

Strong Rejection of Moral Valence a person's guilt cannot alter the moral evaluation of their suffering.

Moderate Rejection of Moral Valence a person's guilt cannot make a person's suffering good.

Tadros argues for the Strong Rejection, in denying that the suffering of the guilty can be good *or less bad*. But the Moderate Rejection allows for the possibility of the permissible punishment of the innocent being *less bad* than the permissible punishment of the guilty, whilst still rejecting the claim that many of us find difficult to accept in retributivism – that the suffering of offenders is good. Moderate Rejection would thus allow us to begin to do justice to the Innocence Intuition.

Let us take stock. Thus far, I have shown that Tadros' substantive view allows for the permissible telishment, framing, and indifference punishing of the innocent in extreme circumstances. I have also

shown that this view, when coupled with Tadros' claim that the suffering of offenders is not good, and is not less bad than that of the innocent, is at odds with the Innocence Intuition. One possibility is to abandon this Strong Rejection of moral valence in favour of Moderate Rejection. Another possibility is to find some other fundamental moral distinction between the harms suffered by offenders and those suffered by innocents. However, such claims must be justified – we need to find a moral difference between the harms suffered in fulfilling the duties of the innocent and those suffered in fulfilling the duties of the guilty.

VI. CHOICES

In the remainder of this article, I want to explore three possible responses to the problems identified above. All require amendments to the Duty View as currently stated. Two, which I find wanting, are developed from Tadros' writings. I develop a third, focusing on rights, which I cannot fully defend here, but which I believe to be more promising. There are doubtless different ways to respond as well.

What could the adherent of the Duty View say in order to differentiate between Sarah and Ruth's punishments? Here is one thing they could say: Sarah *chose* to commit her crime, and thus placed herself in harm's way, while Ruth just ended up in harm's way – it wasn't her choice. However, we should observe at the outset that without further argument this is simply a descriptive difference that distinguishes how Sarah and Ruth came to have their respective duties – it is not yet an evaluative moral difference concerning their having these duties and the harms they suffer in discharging them.

Tadros makes a great deal of the moral significance of choice.[22] But the significance he attaches to it concerns the way in which it can render harm as a means to an end *permissible*. He points to it as *one of* the ways (and, perhaps, the major way) in which we can acquire duties to be harmed, but nothing he says seems to suggest that he views it as a *morally distinctive* way of acquiring duties which would alter the moral evaluations of the harms that lie downstream of those choices (see, especially, p. 58). He says nothing about whether, how, or why we should view harms done to those who

[22] See, for example, pp. 52–58, 169–181, 230–232, 291.

acquire duties in this way differently from harms done to those (like Ruth) who acquire them in other ways, he simply makes it clear that choice is a particularly easy way to justify liability to harm since we had the opportunity to avoid the harm.

However, in attempting to explain the significance of the moral difference between Sarah's situation and that of Ruth, it seems natural to reach for the fact that Sarah acquired her duties through *choice*. Since Tadros thinks the significance of *wrongdoing* and *culpability* in the criminal law is to be found entirely in the significance of choice – 'an account based on choice can fully explain the role that culpability has in determining who is liable to be harmed' (p. 58) – and we are trying to find a distinction between wrongdoers and non-wrongdoers, it is a natural place to search for a moral difference.

Here, however, is a reason to doubt that the significance of choice can fully capture the significance of culpability or wrongdoing. Next to Sarah and Ruth's cells is Beth's cell. A few weeks ago, a terrorist organisation contacted the government and said that unless one innocent person (but it didn't matter who) was punished (by imprisonment for one month), they would commit a terrorist atrocity. Beth was working in the government office at the time the call was taken and volunteered to take the punishment (call this case Beth 1). Now, we may think differently of Beth's predicament than we do of Ruth's, since Beth chose her punishment while Ruth did not. But what is relevant here is that we should think of Beth's situation as importantly morally different from Sarah's, even though both find themselves punished through their choices (Sarah through her choice to murder, Beth through her choice to take the punishment).

Indeed, if we put the significance of wrongdoing entirely down to the significance of choice, we should possibly see Sarah's punishment as *more* troubling than Beth's, for while Sarah only chose to perform conduct that *might* incur a duty to suffer harm (recall that on the Duty View punishment must be *necessary* to avert some future harm, so it is at least theoretically possible that, if there were no further harms or potential crimes to be prevented, Sarah would not acquire any duties to suffer harm), Beth *chose to submit herself to harm* – her suffering is more directly connected to her choice. Another way of putting this is this: Sarah acquires a duty to suffer harm because of her choices; Beth chooses to acquire a duty to suffer harm.

Perhaps it might be objected here that the relevant difference between Beth and Sarah (and indeed Ruth) is that Beth does not actually have a duty to be harmed or punished – she has generously chosen to be punished but has no such duty. However, in volunteering, Beth acquires a duty, like we do when we make promises – we choose to acquire the duty, but it is nevertheless still a duty. Imagine this alternative case (which we can call Beth 2). As in Beth 1, Beth initially volunteers to be punished. After a week or so, however, Beth is tired of being punished and withdraws her consent. However, if we release her now, the terrorists will carry out their threat. We have to keep punishing Beth. And Beth now has a duty to suffer, and to submit herself to continued punishment. Therefore, she has – through her choices – *acquired a duty* to suffer. So, she is in the same situation as Sarah. Yet even though both Sarah and Beth have voluntarily acquired duties to be harmed, their situations, and how we think of the harms that they suffer, are *not* the same. Therefore, I think the significance of choice, whilst important, cannot on its own determine or account for the important moral distinction between the permissible punishment of the innocent and the punishment of the guilty.

VII. THE BADNESS OF HARM VS. CARING ABOUT HARM?

In his responses to other critics Tadros has raised an idea which could help the Duty View to make the distinctions that the Innocence Intuition requires. There, Tadros makes a distinction between the (impersonal) disvalue of some instance of harm, and how much we (from our individual perspectives) *ought to care about* some instance of harm. As an example, my child being harmed is just as serious as your child being harmed from an impersonal point of view. But I ought to *care about* the harm to my child more than I care about the harm to yours. Translated to punishment, Tadros' thought is that offenders' suffering is just as bad (impersonally) as the suffering of the innocent, but we have less reason to care about the suffering of offenders.[23] This would give us a firm line between wrongdoers and the innocent that does not appeal (only) to choice.

[23] Tadros, 'Responses', pp. 257–259.

But this stance raises at least as many questions as it answers. It says that the lives and suffering of offenders should matter less to me, in the same way that your child matters less to me than mine does. However, in the case of our children, my personal relationship with my child gives me clear reasons to care more about my child than yours. In the case of Sarah and Ruth, since (we can imagine) I don't know either of them, it is less clear *why* I should care less about Sarah and her suffering. The only plausible explanation is that Sarah has lost standing of some kind – she used to be just like Ruth in my eyes, and now she isn't. This raises two questions. First, *why* should Sarah have lost this standing – on what grounds should I take it away from her, or regard her as no longer possessing it? It is hard to make a case without appealing to something like desert, the rejection of which animates the entire project of constrained instrumentalism. Second, does this stance really capture the way we view the harms done to offenders? As every bit as morally bad as those done to the innocent, but that we simply *mind* less? This stance gives us *a* line between the punishment of the innocent and the punishment of the guilty, but is it the *right* line? The proposed line doesn't appear to vindicate the Innocence Intuition fully.

VIII. RIGHTS TALK TO THE RESCUE?

In this section I want to explore what I think is the most promising way for the Duty View to accommodate the Innocence Intuition. It has the following virtues. First, it explains, rather than merely postulates, the kind of loss of standing which the response examined in the previous section relies on. Second, it does so without invoking a mysterious separation between the disvalue or moral status of some harm and how much we should care about it – the actual moral status of the harms really do differ. Third, it does so without returning to notions of desert, the rejection of which animates much of the rationale for the Duty View in the first place.

This strategy involves incorporating the concept of rights more fully into the Duty View. There are two important ideas about rights that need to be invoked here. The first is the distinction that some philosophers draw between rights violation and rights infringement. A right is *infringed* when an act implicates the right but is permissible.

A right is *violated* when an act implicates the right and is impermissible.[24] Some theorists think that all punishment implicates rights. That is, we all always have a right not to be punished, but sometimes the state can infringe that right (for example, when it is deserved and/or necessary).[25] If we adopt this line here, however, then we will have no resources with which to make the distinction between the permissible punishment of the innocent and the permissible punishment of the guilty, for both forms of permissible punishment then infringe rights.

Other philosophers do not think that all punishment implicates rights. Rather, they think that, ordinarily, those who may be punished are those who have *forfeited* their right not to be punished.[26] This idea of forfeiture is the second important idea about rights.

We can combine these two ideas: that rights can be forfeited, and that they can be permissibly implicated (infringed). This allows us to make a three-way distinction. Firstly, there are those who may permissibly be punished since they have forfeited their rights (the guilty). (While some, such as Wellman,[27] think that rights forfeiture is a sufficient condition for being punished for any reason, within a constrained instrumentalism framework we would want to say an offender forfeits his rights against being used in certain ways. Therefore only *useful* punishment would be permitted). Then there are those who have not forfeited their rights (the innocent) but who, in extremis, may be permissibly punished. These people have their rights infringed (and therefore will be entitled to compensation if this is possible). Finally, there are those (the innocent in ordinary circumstances) who may not permissibly be punished. These people, if they are punished, have their rights violated.

Thus far, we have some different ways in which it may be permissible to punish people. Let us add that harm that infringes rights is importantly morally different from harm that does not implicate

[24] For a proponent of this distinction, see Joel Feinberg, 'Voluntary Euthanasia and the Inalienable Right to Life' in *Philosophy and Public Affairs* 7 (1978): 93–123. For criticism, see John Oberdiek, 'Lost in Moral Space: on the infrininging/violating distinction and its place in the theory of rights' in *Law and Philosophy* 23 (2004): 325–346; and Alec Walen, 'Innocent Threats vs. Innocent Bystanders: A Case-Study in Rights Theory (unpublished m/s). Walen advocates using claims, in the same way that I consider below.

[25] Husak, *Overcriminalization*, p. 96.

[26] For a recent defence of this view, see Christopher Heath Wellman, 'The Rights Forfeiture Theory of Punishment' in *Ethics* 122 (2012): 371–393.

[27] Ibid., p. 375.

rights at all. In addition, it seems right to add this evaluative claim: all else equal, situations in which rights are infringed are worse than those in which they are not implicated. The advantage of adopting these (independently plausible) positions here is that they allow us to adopt the Moderate Rejection of Moral Valence. Since the suffering of offenders does not implicate rights (since offenders have forfeited rights), it is not *as bad* as the suffering of those whose rights are implicated. We can now say the following. First, there is a fundamental moral difference between the harms to Sarah, who has forfeited her right, and those to Ruth, who has not, and therefore retains her right against punishment, even though both are permissibly punished. Second, all else equal, Ruth's suffering is morally worse, and to be regretted to a greater extent. These two claims allow us to endorse and explain the Innocence Intuition. Third, this account marks the difference between the treatment of Sarah and the treatment of Ruth in a way that captures what, intuitively, is *important* about the difference. When we permissibly punish an innocent person like Ruth, we do not act impermissibly, but we nevertheless act in a way that is morally tainted by the rights infringement.

While they occupy similar territory, rights forfeiture theory and the Duty View are independent theories of punishment. The Duty View seems committed to the idea of rights forfeiture (for guilty persons), as, in acquiring an enforceable duty to submit to punished, they forfeit their right against it. But the forfeiture view doesn't necessarily lead us to a Duty View – it is consistent with my forfeiting a right against punishment that I have no duty to submit to punishment.[28] And it is consistent with the Duty View that innocents who have enforceable duties to submit to punishment *lack* rights against punishment, rather than have their rights infringed. Given this, so far, perhaps what we have is an argument for abandoning the Duty View and switching to rights forfeiture.

But recall why we find duty-based theories compelling in the first place: It seems right to say that Sarah does something wrong when she evades punishment and therefore fails to provide the deterrent she would otherwise provide. And it also seems right to say not only that Ruth may be punished, but that she also has a *duty* to submit to

[28] I am grateful to Kit Wellman for discussion here.

punishment, given that the goods produced by her mild suffering will so vastly outweigh the harms. If, like me, you find both the rights-based and the duty-based stories compelling, is there a way to combine them?

There are three main problems or challenges in trying to tack this three-way view of rights and punishment onto a duty-based view. One is that it requires us to see rights in a certain way, which is itself controversial. The idea of rights infringements requires us to view rights as inputs to deciding what is permissible. They weigh against a course of action, but don't necessarily block it as impermissible. Indeed, Tadros himself argues against seeing rights talk as independent of, rather than following from, what is and is not permissible (p. 201).[29] My own view is that the idea of rights infringement is an important and independently plausible one, but I cannot defend that any further here.

A more serious, second, problem is that to adopt this position on rights alongside the Duty View requires us to endorse the following, very odd statement: Ruth has an enforceable duty to bear a given level of harm, and a right not to be harmed to the same level. It is one thing to say that we can have a right not to be punished and yet can permissibly have punishment forced upon us (though Tadros denies that we should say this, pp. 199–201). But it seems a further conceptual leap to say that I have an *enforceable* duty to Φ and that I have a right not to be forced to Φ.[30] It seems more intuitive to say that, while the innocent person has not *forfeited* her right, she nevertheless *lacks* a right against punishment. However, Jeff McMahan argues convincingly that this is a conceptual possibility: For example, if your duty is to allow yourself to be killed by unjust aggressors in order to save the whole world, he argues, you have a duty to allow this to occur, but also a right against being killed by the aggressors.[31] McMahan's example might avoid the strange conceptual combination of 'right against' and 'duty to' because the right is held against a different group of people to whom the duty is owed. In McMahan's

[29] Elsewhere Tadros has used the concept of rights infringement. See his 'Duty and Liability' in *Utilitas* 24 (2012): 259–277, at pp. 266–277, where Tadros says killing innocent civilians who are not *liable* to be killed may still be permissible but would *infringe* their rights. However, since Ruth has a *duty* to be punished, then Tadros would nevertheless not view her as having a right not be punished.

[30] Tadros himself has written: 'It makes little sense to say that I have an enforceable duty to bear a harm, and yet that I have a right not to be harmed' and that 'we cannot have both an enforceable duty to bear some burden and a right not to bear that burden.' See 'Duty and Liability', pp. 262–263.

[31] Jeff McMahan, 'Individual Liability in War: a response to Fabre, Leveringhaus, and Tadros' in *Utilitas* 24 (2012): 278–299, at p. 295.

example you hold the right against the unjust aggressors, whilst the duty is owed toward the rest of humankind. But imagine that the invaders insisted that your fellow humans (those who will be saved) are the ones who do the killing. This still seems to be a case in which you have a right not to be killed, and a duty to be killed.

Perhaps we can make sense of this if we more carefully specify the duty to be harmed, in the following way: The duty is a duty to *allow your right to be infringed*.[32] That is, you keep your right but you have an enforceable duty to allow others to act contrary to that right. This would mark not only a difference in rights between Ruth and Sarah, but a difference in the content of their duties (as Sarah has forfeited her rights against being harmed, her duty cannot be to allow her right to be infringed).

This may seem conceptually bizarre, but consider the following situation. After Ruth's punishment, the Ruth-haters pack up and leave us alone – their demands have been fulfilled. What should we do with Ruth now? She should surely be compensated – she has suffered (mild) punishment in order to avert harm to others. Yet we wouldn't want to compensate Sarah. Both had a duty to be harmed in this way, but one is entitled to compensation. Allowing that Ruth retained a right against this punishment, whilst Sarah did not, allows us to make sense of the judgement that Ruth can claim compensation while Sarah cannot.

What does this rights-based version of the Duty View allow us to say about Beth, who, recall, volunteers for punishment? The idea that rights may be infringed or forfeited may appear to put Beth on the same side of things as Sarah. Beth, in volunteering to be punished, appears to waive her right not to be punished – in other words, she, like Sarah, appears to lack such a right. This is a third problem for endorsing such an amendment to the Duty View. To provide a separation between Sarah and Beth, we'd need to provide a separation between forfeiting and waiving a right. This could be achieved if, at least sometimes, waiving a right involves allowing others to infringe one's right, rather than removing it altogether.

Perhaps rights infringement is not the best way to describe what happens to Ruth and Beth. An alternative would be to say that, whilst they lack *rights* against being punished (having waived them or by

[32] I am grateful to Mitch Berman for useful discussion here.

simply not having them), they retain a *claim* against being punished. They have (or acquire) duties *not to press this claim*, but it remains there, explaining both the evaluative difference between the harms done to them and Sarah (who has no such claim), and why they should be compensated once pressing their claim becomes permissible. The key point in both of these ways of explaining what happens in these three cases is this. The guilty (Sarah) not only have a duty to be punished, they lack any complaint about being punished, and have no claim to be compensated. The innocent (Beth and Ruth) also have duties to be punished, but there remains some block against their being punished which is overridden by the duty. This block, whether understood as a right or a claim, furnishes us with a moral distinction between the punishments, and with a ground for compensation should the opportunity arise. Crucially, the remaining block against punishment also allows us to make an evaluative distinction between the harms suffered in line with the duties: The harms are less bad when they do not meet rights or claims against the harms.

I cannot here defend all of these controversial positions one would need to endorse in adopting the rights-influenced variant of the Duty View that I have sketched here. I think the duty-based view's problems run deep enough that we must turn to rights forfeiture and infringement (or something like it) in order to explain the difference between punishing the guilty and the innocent. I have tried to explain, however, why the idea of duty remains compelling in this context. Given this, I think we have reason to try to couple the views together.

It is worth outlining how much Tadros, or a supporter of the Duty View, would need to alter their view in order to endorse this position. Since Tadros' view is sometimes described as a rights forfeiture view, and is closely aligned with the self-defence literature in which rights forfeiture is a central concept, it is worth exploring how far we move away from the view if we endorse the duty- and rights forfeiture- based hybrid I have outlined here.[33] First, I think, a key distinction is that, for Tadros, having a duty to suffer harm in the name of harm reduction is supposed to explain *why* the offender lacks a right against punishment. Therefore, the Duty View is supposed to be a deeper story than the rights forfeiture view. But this way of thinking of things would need some finessing if we accept the

[33] I am grateful to a *Law and Philosophy* referee for encouraging me to address this.

view that I have been exploring here, since some who have a duty to suffer harm (the innocent) would *retain* their rights against that harm. So the acquisition of a duty to suffer harm could no longer, on its own, explain the lack of or loss of a right against harm. Second, Tadros is (understandably) resistant to the claim that we can have an *enforceable* duty to Φ and a right against others forcing us to Φ. A perhaps more palatable claim that I have investigated involved having an enforceable duty to Φ and a *claim* against being forced to Φ. But one of these claims would need to be accepted. Finally, Tadros is against a person's moral misdeed's affecting the moral status of their suffering in any way (i.e., he endorses the Strong Rejection of Moral Valence). The position explored here eschews the Strong Rejection in favour of the Moderate Rejection: harms that meet rights (or claims) against that harm are *evaluated* differently from those that do not – they are *less bad*. Another alteration from Tadros' explicit claims is that the innocent sometimes have duties to submit to penal institutions: punishment of the innocent *can* be justified in extremis.

IX. CONCLUSIONS

In this essay I have examined duty-based variants of constrained instrumentalism about punishment, and in particular Victor Tadros' Duty View of punishment, focussing on what they say (or can say) about the punishment of the innocent. Tadros' theory is a philosophically rich one which attempts to find a way to justify the harms of punishment as necessary evils, whilst maintaining tight limits on who will be subject to punishment.

Since these limits on punishment are held by Tadros to be an important element of the theory's attractiveness, it is important to look at this aspect of the theory. Once we do, however, I think the theory, as it stands, has problems. To summarize, above I have tried to show: that the general Duty View takes no stance on if or when the punishment of the innocent is justified – that work is done by claims about *what* duties to be punished the innocent do or don't have; that the specific version of the Duty View which Tadros proposes, and which is most intuitively plausible, actually says that knowingly and intentionally punishing the innocent can be justified, despite Tadros' various statements to the contrary; and, most

importantly, that the theory at present cannot morally distinguish between permissible punishment of the innocent and permissible punishment of the guilty.

In rejecting any role for desert, constrained instrumentalism views rob us of one of the simplest and clearest ways of maintaining an all-the-way-down, fundamental distinction between the innocent and the guilty. Whilst I think Tadros' theory is a brilliant achievement – a genuinely original contribution to penal philosophy – unless it can make this distinction I, for one, will find it difficult to endorse. I have tried to show that bringing rights (or claims) forfeiture and infringement to the heart of the theory may be a promising avenue, although doing so may require endorsing controversial positions on rights – not least the idea that one can simultaneously hold a right against being punished, and a duty to submit to punishment.

ACKNOWLEDGEMENTS

I have benefitted from discussion with, and/or comments from, Mitch Berman, Antony Duff, James Edwards, Kim Ferzan, Tom Parr, Massismo Renzo, Adam Slavny, François Tanguay-Renaud, Malcolm Thorburn, Alec Walen, and Kit Wellman. The paper was presented at a workshop on Tadros' book in Split, Crotia, and at a 'New Voices in Criminal Law Theory' event in Death Valley, California. I am grateful to both groups for stimulating discussion, and in particular to Doug Husak who commented on the paper in Death Valley. The paper was further improved in response to comments from two referees for this journal, in particular Alec Walen. My greatest debt is to Victor Tadros: for providing the theory with which the present paper engages; for comments on several draft versions of the present paper; and for many conversations about criminal law theory and moral and political philosophy more generally. Of course, in philosophy it is criticism, not imitation, that is the sincerest form of flattery.

OPEN ACCESS

Department of Politics and International Relations,
University of Reading, HumSS 405,
Whiteknights Campus, Reading, RG6 6AA, UK
E-mail: p.r.tomlin@reading.ac.uk