Two-Tiered Mixed Theories of Punishment Are Not Safe from the Angry Mob

Jason Lee Byas

University of Michigan

Abstract: Two-tiered mixed theories of punishment hold that legislatures should act according to consequentialism, but the judiciary should act according to retributivism. A major motivation for these theories is wanting to preserve the idea that punishment is ultimately justified on consequentialist grounds, without falling prey to the Punishing the Innocent objection. Yet this benefit is illusory. While two-tiered mixed theories successfully avoid the Punishing the Innocent objection narrowly construed, they do not successfully escape the point behind it. This is because cases can be constructed with the same problems as those used to describe the classic Punishing the Innocent objection, just bumped up to the legislature.

(Forthcoming in *Australasian Journal of Philosophy*.)

1 Introduction

The most common objection to consequentialist theories of punishment is that they sometimes require punishing the innocent. The classic case used in this objection is one in which a mob threatens to wreak havoc unless an innocent person is convicted and punished. It appears that if consequentialism is true, we are obligated to acquiesce. Partly to avoid this issue, some who are otherwise attracted to penal consequentialism have developed mixed theories of punishment with a two-tiered structure. On these views, legislatures should craft laws and punishments according to consequentialism, but judges and juries should only convict and sentence based on guilt.[1] The idea is that

[1] At various points in this paper, I refer to how 'judges and juries' should 'convict and sentence.' To be clear, while there is variation between jurisdictions, it is typically the case that juries decide whether to convict and judges decide how to sentence. An exception in the United States is that juries impose sentences when the death penalty is the possibility. Regardless, two-tiered mixed theories of the kind under consideration

since judges and juries are deliberating as strict retributivists, the temptation will not arise, and so we can rely on consequentialism to design the criminal law without punishing the innocent.

I argue here that to whatever extent punishing the innocent is a problem for penal consequentialism, the point behind that objection is also a problem for these mixed theories. This is true even if we grant that such theories can consistently hold to both legislative consequentialism and judicial retributivism. This is because the same problems can get bumped up to the legislative stage. Whatever response the mixed theorist could give to this new case, the penal consequentialist can give that same response in the original case. In other words: if these problems are fatal to penal consequentialism, they are fatal to mixed theories; if mixed theories survive the challenge, so does penal consequentialism.

I begin by briefly reviewing the Punishing the Innocent objection and why it is a problem for penal consequentialism. Then in section three, I give a rough outline of two-tiered mixed theories and how they are meant to avoid punishing the innocent. In section four, I give two cases that are structurally similar to the Angry Mob case, except that they involve a choice by legislators rather than judges or juries. In that same section, I highlight how those features that are objectionable in the original case are once again present in the legislative versions. So too do I note that some responses potentially available to the mixed theorist mirror responses straightforward consequentialists have given to the original case. Section five considers and rejects the possible reply that my argument misunderstands the force of the Punishing the Innocent objection by focusing too heavily on the Angry Mob case used to explain it. In the sixth section, I consider one further response, which is to extend the mixed theory's deontological components to the legislative process.

2 Penal Consequentialism and Punishing the Innocent

By 'penal consequentialism,' I refer to theories in which punishment is justified entirely by its consequences, be those the preservation of public order, general happiness, or whatever else. This may be connected to a broader consequentialist moral theory, but it need not. When it is not connected to consequentialism as a broader moral theory, it is likely to be more specific about the consequences that punishment is after, such as personal security or public order, rather than utility.

---

here suggest that both judges and juries should deliberate as if they are strict retributivists.

Any such view will be met with the objection that it sometimes demands of judges and juries to punish people who they themselves know to be innocent. Philippa Foot's statement of the case is representative:

> Suppose that a judge or magistrate is faced with rioters demanding that a culprit be found for a certain crime and threatening otherwise to take their own bloody revenge on a particular section of the community. The real culprit being unknown, the judge sees himself as able to prevent the bloodshed only by framing some innocent person and having him executed' (1967: 2).

Assuming that the mob cannot be placated or overpowered, and that the real culprit will not be found soon, it seems that we must either kill one innocent person or allow many more to die. If punishment is justified by its social consequences, those consequences are likely to have something to do with preventing harm to other people. While punishing an innocent person here will itself be a serious harm, much worse harm would come from refusing to do so. Thus, if social consequences themselves are enough to justify punishment, it looks like punishing the innocent would in this case be justified. This strikes most people as false, and so it is a persistent objection to the consequentialist reasoning that got us there.

For our purposes, it is helpful to review both why the apparent consequentialist conclusion seems so perverse and some common consequentialist replies.

A first problem with punishing the innocent in this case is that it sacrifices someone for others' benefit, which treats the sacrificed person as a mere means. A second is that the cost-benefit judgment here is what it is as a function of the rioters' bloodlust. Something feels off about *caving to the mob* being elevated to the level of moral principle. Third, this is especially so in the legal context: punishing the innocent is in part perverse because it subverts the very purpose of the criminal law, which is to respond to certain kinds of wrongdoing. The law is meant to respond to murder, and here it is being used to commit it.

This last problem also leads many consequentialists to say that if punishment is an institution worth establishing for consequentialist reasons, the nature of the institution will forestall directly applying consequentialism to individual convictions (Quinton 1954; Rawls 1955). This sort of response is to abandon straightforward penal consequentialism for the kind of mixed theory at issue in this paper, so we can put it aside for now.

More straightforward penal consequentialists generally follow one of two lines suggested by Kai Nielsen. First, to deny that the apparent conclusion actually follows from even a direct application of consequentialism in this case. For instance, Nielsen argues that word would eventually get out about the fact that the punished person was not actually guilty. This would corrupt the legal system: no one would trust verdicts,

mobs would know they could get what they want through threats, and less scrupulous judges would be emboldened for their own false judgments in other cases. So, the consequentialist might argue that their theory argues against caving to the mob. Even though it allows more harm than the immediate harms to the innocent person punished, there is more harm at risk than that involved in the sacrifice itself (1972: 223–4).

A second response is to say that biting the bullet here might not be so bad. If we judge the apparent consequentialist conclusion against the thought that the innocent should not be punished *no matter the consequences*, its perversity is not so obvious. If our options were between punishing the innocent and allowing the destruction of the universe, for example, it is hard to argue for the latter. To the extent that empirical facts will not allow for the response of the preceding paragraph, the consequentialist can argue that refusing to punish the innocent must surely sometimes be justified (1972: 224 –7).

Whatever we might say about the plausibility of these responses, they are ways that the straightforward consequentialist might go. It will be helpful to keep them in mind when considering the mixed theorists' potential responses to the cases in section four.

3 Two-Tiered Mixed Theories of Punishment

Faced with the Punishing the Innocent objection and other worries, many philosophers with consequentialist inclinations about punishment opt instead for *two-tiered mixed theories*. These are views in which legislators should use consequentialist reasoning in deciding what to punish and how, but judges and juries must convict and sentence on retributivist grounds. By separating which justifications are acceptable at the legislative stage from those acceptable at the judicial stage, such a theory follows a *two-tiered* approach.

Many theorists follow this route, and among them there are many different explanations for separating how punishments should be justified by legislatures and judiciaries. Some, such as Anthony Quinton, argue that the very definition of punishment includes that it is a response to lawbreaking. Accordingly, we are prevented from 'punishing the innocent' on purely logical grounds, even if the justification of punishment is consequentialist (1954). Others, following Rawls, hold that once a *practice* (such as punishment) has been justified, having the practice at all requires acting within it on its own terms, not external ones such as overall utility (1955). The most famous mixed account is that of H. L. A. Hart, who splits the question of punishment's justifiability into that of its 'General Justifying Aim,' which asks 'why have punishment at all?,' and its 'distribution,' which asks 'once we have punishment, who should be punished' (1960)?

Importantly, while some mixed theories find their justification in rule consequentialism, mixed theories are not equivalent to rule consequentialism, nor need they rely on rule

consequentialism.[2] Matthew Altman has recently defended a sophisticated two-tiered mixed theory in which this is especially evident. He argues that the state has many different obligations, and these include both preserving the public order and expressing the public's resentment towards wrongdoing (2021). In other words, both penal consequentialism and expressive retributivism are true. Yet different parts of the state have different epistemic abilities. In particular: the legislature is best equipped to consider wide-ranging consequences of policies on the public order, whereas the judiciary is best equipped to consider the precise guilt of particular people in particular cases. Since ought implies can, the extent to which we are obligated by different moral concerns is partially shaped by our epistemic abilities (2021: 121–31). Thus, the state's consequentialist obligation to preserve the public order has its strongest hold over the legislature, while its retributivist obligation to express rightful public resentment is most relevant for the judiciary.[3]

In any event, these mixed theories more or less share the basic thought that legislators should act in accordance with consequentialism when crafting laws and setting a range of punishments, whereas judges and juries should act in accordance with retributivism when settling particular cases. For this reason, mixed theorists are less bothered by those cases generally used to raise the Punishing the Innocent objection. On the mixed theory, judges should not be thinking in consequentialist terms at all, despite the fact the punishments they inflict are ultimately justified by their consequences. Thus, even when confronted with an angry mob, judges following a mixed theory of punishment will refuse to punish the innocent.

4 The Angry Mob Invades the Legislature

Some argue that mixed theories fail because they are hopelessly ad hoc (Kaufman 2008; 2013: ch. 4). Others doubt that such theories can really get away from the conclusion

---

[2] Here and elsewhere in the paper I use the phrase 'mixed theories' to refer specifically to two-tiered mixed theories. To be clear, there are many other ways one might mix retributivism, consequentialism and other theories of punishment that do not fit this two-tiered approach of legislative consequentialism and judicial retributivism. My criticisms here are meant only for those theories that do take this approach. My use of 'mixed theories' in this way is only for ease of exposition. See Hoskins (2021) for a discussion of the broader landscape of mixed theories.

[3] This epistemic argument is only one of many arguments Altman gives for trusting the legislature with the state's consequentialist obligations and the judiciary with the state's retributive obligations. For instance, he also rejects following consequentialism at the judicial level because of the classic Punishing the Innocent objection discussed here (2021: 148). See also Altman (2023) for still further arguments.

that judges should sometimes knowingly inflict suffering on the legally innocent (Boonin 2008: 63–77). Another critique holds that retributivism and consequentialism rely on incompatible understandings of offenders and their responsibility (Lippke 2006). I set aside these worries, because even if the two tiers are stably separable in a principled way, and even if it is coherent to jointly endorse retributivism and consequentialism, there is another problem. This is that the preservation of penal consequentialism at the legislative stage allows for scenarios quite similar to the classic Angry Mob case, just bumped up to the legislature.

Here are two.

Political Necessity Suppose the legislature is controlled by a political party with what you take to be a perfect penal policy, which best achieves those consequences you think punishment should pursue. Yet they are not the only party, and the opposition is miserably bad. The party out of power seeks to criminalize many things that have positive or neutral consequences. Not only that, they also aim to replace existing punishments with counterproductively brutal ones. This might be the death penalty, life without parole in intentionally inhumane conditions, solitary confinement, public torture, or whatever other punishment you think fits this bill. Riding on a crazed populist wave, this opposition seriously threatens to take back the legislature in a landslide. Dependable polling also shows that the current establishment can rest assured they will secure power, and so prevent unjust prohibitions, but only if they install some of those counterproductively brutal punishments for various existing crimes.

Safer Sacrifice Assume that blasphemy laws are unjust. Suppose further that the call for blasphemy laws is quite strong, and without any on the books, vigilantes will enforce such a prohibition, and it will be quite brutal. Both the military and police are ill-equipped to stop this campaign of violence, and many in their ranks are even sympathetic to it. Yet if the legislature just imposes blasphemy laws of their own, enforced mostly by fines, this will satisfy enough would-be vigilantes that the threat will disappear.[4]

If legislators are acting in accordance with penal consequentialism, it seems that counterproductively brutal punishments must be installed in the first case, and blasphemy laws must be imposed in the second. On the mixed theorist's own grounds,

---

[4] I give two cases here for two reasons. First, because a mixed theorist might allege that the technical details of one case or the other make it inapplicable to their theory. Second, to demonstrate that once the basic problem is made clear, it is trivially easy to construct cases of this kind, just as it is trivially easy to construct cases where it seems that the unmixed penal consequentialist must punish the innocent.

then, they must impose what they would otherwise see as unjustified legal practices in the first case, and unjustified legal content in the second.

This is because in both cases, the sorts of consequences mixed theorists want to consider when deciding what to punish and how will go in a significantly worse direction if legislators do not so-act. If the opposition party comes into power, more people's rights will be violated, and there will be a significant decline in public safety. So too if the anti-blasphemy vigilantes continue their campaign.[5]

These conclusions mirror the straightforward consequentialist's conclusions that raise the classic Punishing the Innocent objection, and so the mixed theorist does not really avoid the point behind the objection. Punishing the innocent is unacceptable because it involves sacrificing people as a mere means for others. Legislators imposing horrific punishments in Political Necessity do this by killing, torturing, or doing whatever else you might fear to some simply because it prevents other cruel punishments from falling on others. So too does a legislature who criminalizes blasphemy in the Safer Sacrifice case harm peaceful people simply to prevent further damage.

Punishing the innocent was also unacceptable because it held morality hostage to the unjust threats of an unreasonable mob. This, too, is true of the legislative analogs. In Political Necessity, the populist wave fills the role of the mob. In Safer Sacrifice, laws adjust to meet the demands of more literal mobs.

Additionally, there is a perversity behind punishing the innocent in that it subverts the purpose of the institution. Mixed theorists themselves would tell us that the legislature's purpose in establishing the criminal law is something like the protection of the public. There is of course a trivial sense in which, in order for my point to work, this is fulfilled in both the Political Necessity and Safer Sacrifice case. Yet this is quite a distorted form of 'protecting the public': deliberately taking actions that enact cruelty or restrict liberty with significant further costs to security and no benefits, *except* in that doing so removes threats made to force those same policies.

Mixed theorists have some replies available. For example, it may be that in both cases, the morally horrifying course of action is not actually the one that best satisfies the legislature's consequentialist obligations. There may be some other way to win the election, or to prevent the opposition party from doing too much damage once they get into power. So too might the right choice just be to crack down on the anti-blasphemy vigilantes despite the odds. Yet so too could the straightforward consequentialist insist

---

[5] I emphasize these harms because many mixed theories are not guided by utility as-such in their consequentialist modes, but instead something more specific. Altman, for example, focuses on 'preserving the public order' (2021: Ch. 2). Other theories are likely to focus on something in that neighborhood, so the harms in my example are specifically ones involving threats to public order.

in response to the classic case that maybe police will find the real culprit any minute now, or that really we should focus on quelling this angry mob. To the extent that these are unsatisfying replies from the penal consequentialist, so too will the above be unsatisfying from the mixed theorist.

Alternatively, the mixed theorist could appeal to longer-range consequences. In the Political Necessity case, it may be that instituting cruel punishments will win you this election at the cost of corrupting the party. That it was a cynical political move will be obvious to everyone, alienating your party's strongest activists and emboldening the opposition to push for even worse compromises. Similar arguments could be made in response to the Safer Sacrifice case, especially in terms of emboldening vigilantes. Yet so too could similar arguments be made in the original Punishing the Innocent objection, as these mirror Nielsen's first sort of reply.

Biting the bullet might also be tempting. Maybe it's not so bad to make a political compromise that stops a horrifying slew of policies, and maybe it's not so bad to punish something with a fine to prevent vigilante murders. Yet biting the bullet in this same way is also available to the consequentialist in the original case.

In general, there are things that the mixed theorist can say, but they do not put them in much better of a position than the straightforward consequentialist. Even if the mixed theorist can evade the classic Punishing the Innocent objection as it is traditionally presented, they do not escape the real problems that the objection is meant to highlight.

5 Leaving Aside the Angry Mob

In reply to these worries, a mixed theorist might agree about the symmetry between their replies to Political Necessity and Safer Sacrifice and the penal consequentialist's replies to the Angry Mob. Yet it does not necessarily follow that this symmetry is symmetry with respect to the Punishing the Innocent objection. This is because the Angry Mob is a particular case meant to show the force of the Punishing the Innocent objection, but it is not that objection itself. There are many other kinds of scenarios, with different structures, in which it seems the penal consequentialist would have to punish the innocent, and it may be that these scenarios do not have close parallels for the mixed theorist. Moreover, it is still the case that the mixed theorist need not literally punish the legally innocent under that exact description in many of the cases where the more straightforward consequentialist must, such as with the original Angry Mob. Perhaps, the mixed theorist might argue, these differences are significant enough to merit rejecting consequentialism for reasons that do not apply to mixed theories.

This reply fails for three reasons. The first two emphasize that even if mixed theories need not require punishing the legally innocent, they would still run afoul of the reasons punishing the legally innocent strikes us as so perverse. More to the point, the third

problem is that certain variants of the cases above *will* require us to knowingly punish the legally innocent.

First, intuitively, the reason that punishing the innocent is so bad is that it involves things like sacrificing individuals or subverting the purpose of the law.[6] It is very hard to see what the force of the objection could be if it is very clearly separated from factors like this. Perhaps it is that there seems to be a contradiction in using the law to punish against the standards laid down by the law. Yet if this is so, then the force of the problem is something like subverting the purpose of the law. If it is the case that mixed theories also face the problems behind the Punishing the Innocent objection, then more needs to be said about why the specific way in which consequentialism faces the problem is significantly worse.

So too must it be clear that the specific problem in question could not be easily replicated. For example, it might be thought that at least in the legislative cases, these are laws, and so those subject to them have fair warning in a way that is not true of a person who just finds themselves charged for a crime they did not commit. Yet if the legislature is totally guided by consequentialist considerations in constructing the criminal law, one could easily imagine a version of either Safer Sacrifice or Political Necessity in which legislation is rushed through and immediately enforced, or even in which they are written to apply ex-post-facto. In either event, those who face the brunt of the new law would not have fair warning, and so the fair warning of law cannot be what gives us the principled distinction that mixed theorists need.[7]

Relatedly, the second general problem for this reply of pulling the Punishing the Innocent objection apart from the Angry Mob example is that most other cases where consequentialism demands punishing the innocent have ready mirrors for mixed theories. For example, it may be that a judge or prosecutor is concerned with their future electoral success, because they know that they would do a better job of protecting the public than their opponents. This would be even closer to the Political Necessity case than the Angry Mob, so the Political Necessity case already serves as a ready analog.

It might instead be that virtually everyone believes that someone accused in a high-profile case is guilty, even though they are in fact innocent, as understood by judge and jury. Suppose that while there is no angry mob issue to consider here, there is reason to think the case will have a significant effect on deterrence, given its high profile. Here the consequentialist would need to punish the innocent to achieve good consequences, not

---

[6] Recall my discussion of these points in section two.

[7] This reply may be unsatisfactory for some readers. If so, then what I say in my third reply below may help, seeing as there even the safety of legal innocence will be taken away, and so too will its fair warning.

to avoid bad ones. Yet similar problems crop up for the mixed theorist at the legislative stage: suppose it were so that people falsely but *very, very strongly* associated some legitimate sector of the economy with criminality, thinking they are fronts for organized crime. Suppose further that this association is so strong that the presence of that kind of business in a community leads people to think crime is tolerated. It seems, then, that a legislature operating on consequentialist principles could have a reason to ban this kind of business, even if they know that the stereotype is baseless, since banning it could have positive deterrent effects.[8]

The point in these cases, as with Political Necessity and Safer Sacrifice, is not that the specific case is a problem. It is that they draw out a much more *general* problem with mixed theories' guidance to set laws and punishments in a consequentialist way, and that these problems mirror those with full-blown penal consequentialism's guidance for conviction and sentencing. Importantly, once the basic problem is made clear, cases illustrating it are *trivially easy to construct*. This gives us reason to think that the general features of judicial consequentialism that raise Punishing the Innocent worries are general features of allowing consequences to sufficiently justify punishment, even when that 'general justifying aim' is constrained to guiding the legislature. It is very unlikely, then, that the Punishing the Innocent objection will be a good reason to prefer mixed theories over thoroughgoing penal consequentialism, because it appears that any case used to show the force of the Punishing the Innocent objection will have a clear legislative analog.

Unless, of course, the mixed theorist can more clearly draw out what is significantly wrong about punishing the innocent in a way that does not involve legislatively replicable factors like sacrificing individuals, caving to mobs, or subverting the purpose of the law. Even if they do, however, there is still a third problem. This is that there are some scenarios in which a two-tiered mixed theory might demand that the legislature act to punish the legally innocent, without having the legislature directly convict or sentence criminals.

Arguably, further background information could make Safer Sacrifice such a case. Suppose that the constitution of the country in question has it that there can be no laws restricting religious speech of any kind. Any prohibition on blasphemy would violate this, and so be unconstitutional. There is a sense, then, in which those prosecuted under

---

[8] Additionally, cases of avoiding bad consequences can typically be easily flipped to one of achieving good consequences. For instance, suppose that the more humane party in Political Necessity is out of power, and can win the election by knowingly opting for new, counterproductively brutal punishments, while also greatly reforming other aspects of the criminal law.

the blasphemy law in this modified version of Safer Sacrifice are legally innocent, because the law in which they are being prosecuted is itself illegal.[9]

However, it might sound strained to some to describe those who are unconstitutionally prosecuted as 'legally innocent.' Thus, it is worth reiterating an even clearer kind of case in which a legislature might act to punish the innocent on consequentialist grounds, without directly involving itself in conviction or sentencing. This is Saul Smilansky's suggestion that utilitarianism might require knowingly punishing the innocent at a more systemic level, through lowering standards of evidence to suppress crime (1990).

Of course, setting standards of evidence in the real world will necessarily be a difficult balance between those that insufficiently convict the guilty and those that excessively convict the innocent. As Smilansky observes, though, the utilitarian method of making that tradeoff through a simple cost-benefit analysis treats the risk of false convictions much more flippantly than any method guided by non-consequentialist respect for innocents' dignity (1990: 230). Given some plausible empirical conditions, utilitarianism would require us to dramatically relax evidentiary standards and knowingly punish the innocent at a much greater rate than we do now (1990: 259). What is important for our purposes is that any such changes to evidentiary standards are likeliest to come through the legislature, and so the point would seem to apply just as well to the two-tiered mixed theories under consideration here.[10]

Altman suggests briefly that while Smilansky's point may apply to classical utilitarianism, 'it does not impact consequentialism more generally, especially a consequentialism of rights' (2021: 103). If so, then perhaps it does not impact more sophisticated two-tiered mixed theories like his, which may appeal to a consequentialism of rights.

I am not sure that this is actually true with respect to Smilansky's original version of the problem. Notably, Smilansky's points rely on the relaxed standards creating greater deterrence and incapacitation, and thus reducing an innocent person's chance of being victimized by crime, which may well be much higher than their chance of being wrongly convicted (1990: 259). If Smilansky is right about this to begin with, then it seems that a

---

[9] In the interest of emphasizing that these cases are trivially easy to construct once their basic structure is made clear, I invite the reader to imagine a modification of Political Necessity that would require a legislature guided by consequentialism to violate their nation's constitution.

[10] Notice that since the legally innocent are now at greater risk of punishment here and in the variation considered below, the 'fair warning of law' advantage considered earlier in this section disappears.

'consequentialism of rights' would prefer the lesser risk of one's rights being violated in wrongful conviction to the higher risk of one's rights being violated in crime.

Nonetheless, even if Smilansky's deterrence-based version of the point did not hold for mixed theories where legislatures should act on 'a consequentialism of rights,' it would be simple enough to adjust the Political Necessity case for a variation of his conclusion.

Suppose that one of the politically popular but criminologically mistaken reforms demanded by the opposition party is to greatly lower the burden of proof for certain crimes. Suppose further that the party in power knows well that lowering the burden of proof in the relevant way will have no added deterrence effect, but will greatly increase the number of innocent people convicted for those crimes. Finally, suppose most unfortunately that dependable polling suggests the party in power must enact this policy to win. Given the stakes already described in Political Necessity, the alternative is to accept much worse changes to the law that will involve counterproductively brutal punishments for things that should not even be criminalized. Thus, it seems that a legislature choosing policies with an eye to rights consequentialism should counterproductively change the burden of proof in a way that will knowingly lead to a sharp increase in the number of innocent people punished.[11]

Of course, the same caveat applies to the cases mentioned in this section that applied to the original versions of Political Necessity and Safer Sacrifice: there are plenty of things the two-tiered mixed theorist can say in reply to these objections, so what I have said here is far from conclusive. Yet the things they can say will mirror what pure penal consequentialists can say in reply to the original Angry Mob case. Thus, the two-tiered mixed theorist's fate with respect to the point behind the Punishing the Innocent objection will be identical to the consequentialist's. Any reason to favor two-tiered mixed theories over penal consequentialism must be found elsewhere.

6 Restricting the Legislature by Adjusting the Mixture

So far, I have assumed that legislatures following two-tiered mixed theories must act in a purely consequentialist way, just as judiciaries following two-tiered mixed theories must act in a purely retributivist way. However, many proponents, such as Altman, will accept both the deontological and consequentialist parts of their theory as genuinely normative, with neither as a mere means to the other. One way to handle the problems discussed in the last two sections, then, is to say that the deontological elements of the theory should restrain the legislature's pursuit of consequentialist goals. This might be by adopting a 'negative retributivism,' in which retributive desert is required for the

---

[11] Incidentally, the variant of Political Necessity I give here also serves to further Smilansky's original critique of straightforwardly utilitarian theories of punishment, as it shows that his point will apply even in cases where actual deterrence is not at issue.

permissibility of punishment, but does not give us positive reason to punish. Another way this might go would be to accept some form of rights forfeiture, in which those who commit crimes forfeit their rights against being punished.

Here the thought would be one of the following two options: first, that the legislature may only punish those things for which people accrue preinstitutional retributive desert, and must only punish them within the bounds of that desert. Given those constraints, they must further only criminalize those things which it would aid consequentialist goals on net to criminalize, and punish them in the way that would best satisfy those goals. Alternatively, only those activities that forfeit rights are open to being criminalized, and legislators should construct punishments such that they will only act against rights that offenders have forfeited. Here again, they would pursue consequentialist goals within these constraints.[12]

Something Altman says very briefly in a different context might suggest this reply with respect to rights forfeiture:

> …In a liberal political system …, the purpose of the law is to protect the rights of its people, and any power the government has derives from the consent of the governed. … [A constitutional] prohibition on "cruel and unusual punishment" recognizes that even convicted criminals have some fundamental rights that should not be violated; we owe them certain protections even though they have forfeited some of their rights in committing crimes. (2021: 197)

Similarly, just before this, he says

> We should avoid punishments that are deserved but are inconsistent with a society's values. And if the punishment is wrong in itself or the system it supports is immoral, then it should not be carried out, regardless of how effectively it controls crime (2021: 196–7).[13]

The meaning of these suggestions is unclear, at least in the context of Altman's broader framework. Altman discusses rights forfeiture or unforfeited rights nowhere else in the book, other than in discussing alternatives to his view (2021: 68–70). Additionally, he does not elaborate on what it means for a punishment to be wrong in itself or part of an independently immoral system. Given that Altman takes consequentialism about preserving the public order to be what justifies punishment in the first place, answering

---

[12] For more on negative retributivism, see Lippke (2015). For more on rights forfeiture theories of punishment, see Wellman (2017).

[13] Altman's own context for both of these suggestions is the importance of law in constraining the criminal legal system (2021: 196–7), before then going to give a novel defense of jury nullification (2021: Ch. 11).

the abolitionist thought that *all* of criminal punishment is wrong in itself or part of an immoral system, we would need to hear more about what might set aside some punishments as wrong irrespective of consequences to the public order. Otherwise, it seems that if something is good enough to overcome strong moral presumptions like freedom of movement and rights to personal property, it would be enough to overcome other strong moral presumptions.

However, Altman does briefly explicitly affirm negative retributivism as a part of the two-tiered approach:

> …[R]etributivism under the two-tiered model is both negative and positive, but at different levels. Considered from outside the model, as it were, desert is only a necessary condition for punishing an offender. It is not a positive obligation, since an offender's guilt does not sufficiently justify legal punishment, in the absence of applicable laws, and criminal statutes are justified on consequentialist grounds. However, judges, who apply the laws to specific criminals, are positive retributivists, in the sense that actual guilt is sufficient to find them guilty and to punish them in accordance with their degrees of guilt, within the statutorily defined punishment schedule. Another way of putting this is that I, as a defender of the two-tiered model of punishment, am a negative retributivist, but judges, as agents who are bound by the theory's constraints, are positive retributivists (2021: 200).

If the two-tiered approach could be amended with negative retributivism, this would greatly reduce and perhaps even resolve the worries I have raised. Yet this is all that Altman says about negative retributivism as a constraint on the punishments legislatures might construct.[14] We would need much more to see if this mixture really worked as promised. Pending that, we have two reasons for skepticism, given the dialectic of Altman's argument and the kinds of reasons that attract philosophers to two-tiered mixed theories.

First, we now need a theory of preinstitutional retributive desert or rights forfeiture, and it must be of a kind that can be roughly measured prior to crafting the criminal law and setting punishments. If we have neither, and concepts like desert and rights forfeiture only come about in virtue of the laws laid down by the legislature, then desert and unforfeited rights cannot restrain the legislature. If preinstitutional desert and

---

[14] In fact, it is somewhat ambiguous as to whether Altman thinks this should serve as a constraint on the legislature in their own deliberation. The quoted passage is from his defense of jury nullification, arguing that juries should nullify when laws are patently unjust (2021: Ch. 11).

rights forfeiture are too epistemically difficult to assess, then they will not be of much help in knowing the legislature's limits.

This is especially problematic for the mixed theorist, since it is often dissatisfaction with ideas like preinstitutional retributive desert or their application to the law that partly motivates their legislative consequentialism (Altman 2021: 50–53; 125; Ch. 10).[15] For example, the mixed theorist cannot coherently hold that preinstitutional desert is impossible to measure when arguing against traditional retributivism only to then appeal to that same preinstitutional desert in answering worries for legislative consequentialism. If it is impossible to measure preinstitutional desert, then we will have no way of knowing how far is too far when the legislature is deciding which punishments to attach to vandalism in the name of preserving public order. If it possible to measure that preinstitutional desert when deciding that, then it is also possible for straightforward desert retributivists to measure it when constructing laws on retributivist grounds.[16]

Second, if the two-tiered mixed theorist also relies upon negative retributivism or rights forfeiture at the legislative level, it becomes harder to see what separates them from the straightforward retributivist or rights forfeiture theorist. After all, retributivists and rights forfeiture theorists also think that consequences can *matter*, and that fact alone is not enough to make them mixed theorists.[17] On this hypothetical, modified version of the two-tiered mixed theory currently being considered, it is the deontological elements that ultimately justify the permissibility of punishment in the first place, and then again the deontological elements that justify the particular punishment of a particular person. All that the consequentialist elements do is tell the legislature when punishment is all

---

[15] The most relevant of these citations to Altman is his Chapter 10, which is a sustained attack on the idea that proportionality is best set with a mind to retribution. If this is so, then retributive proportionality also cannot serve as a negative constraint.

[16] There is a similarity here between this dilemma and that of answering Political Necessity and Safer Sacrifice without letting straightforward consequentialists off the hook for the Angry Mob. In both cases, the mixed theorist cannot have their cake and eat it too when rejecting the pure version of a theory while also helping themselves to that theory's resources without showing why those resources are still available.

[17] Zachary Hoskins argues otherwise, holding that rights forfeiture and negative retributivism must be components of a mixed theory, rather than fully-fledged theories of their own (2021: 39; 46). However, this is typically not the self-understanding of negative retributivists or rights forfeiture theorists. For this with respect to the former, see Lippke (2006) in conjunction with Lippke (2015), with respect to the latter, see Wellman (2017). Moreover, Hoskins's discussion of mixed theories is explicitly much broader than my narrow focus on those with a two-tiered structure.

things considered worth doing and remind them to make sure it gets doled out productively. It is unclear why the straightforward retributivist or rights forfeiture theorist would see this as a competitor to their approaches, rather than a version of it.[18] Much more needs to be said to explain how the mixed theorist can appeal to deontological parts of their theory at the legislative stage without collapsing into an unmixed retributivist or rights forfeiture theory.

7 Conclusion

Here I have argued that even if we grant mixed theorists of punishment the claim that they avoid the classic Punishing the Innocent objection as it is typically described, they do not ultimately fare much better than full-fledged consequentialists on the problems behind that objection. This is because the same structure of the traditional case can be replicated at the legislative stage. Those who favor full-fledged consequentialism have reason to welcome this conclusion, as do retributivists, rights forfeiture theorists, penal abolitionists, and others who reject mixed theories. For the mixed theorist hoping to avoid this conclusion, they must find compelling answers to cases like those above that cannot be given by straightforward consequentialists to the classic Punishing the Innocent objection, and which do not involve abandoning their mixed approach for more some robustly deontological alternative. Otherwise, they cannot appeal to the Punishing the Innocent objection as a reason to prefer their view to straightforward consequentialism.

References

Altman, Matthew C (2021) *A Theory of Legal Punishment: Deterrence, Retribution, and the Aims of the State*. Routledge.

Altman, Matthew C (2023) 'In Defense of a Mixed Theory of Punishment', in Matthew C Altman, ed., *The Palgrave Handbook on the Philosophy of Punishment*: 195–219. Palgrave Macmillan.

Boonin, David (2008) *The Problem of Punishment*. Cambridge University Press.

Foot, Philippa (1967) 'The Problem of Abortion and the Doctrine of the Double Effect', *Oxford Review* **5**: 5–15.

Hart, H L A (1960) 'Prolegomenon to the Principles of Punishment', *Proceedings of the Aristotelian Society* **60**: 1–26. doi:10.1093/aristotelian/60.1.1.

Hoskins, Zachary (2021) 'Hybrid Theories of Punishment', in Farah Focquaert, Elizabeth Shaw and Bruce N Waller, eds., *The Routledge Handbook of the Philosophy and Science of Punishment*: 37–48. Routledge.

[18] Even an especially robust retributivist like Michael Moore holds that desert can be outweighed by other considerations, and so agrees that consequences matter in how we construct the criminal law (Moore 1997: 173–5; 2016: 344–7).

Kaufman, Whitley (2008) 'The Rise and Fall of the Mixed Theory of Punishment', *International Journal of Applied Philosophy* **22**: 37–57. doi:10.5840/ijap20082214.

Kaufman, Whitley R P (2013) *Honor and Revenge: A Theory of Punishment*. Springer.

Lippke, Richard L (2006) 'Mixed Theories of Punishment and Mixed Offenders: Some Unresolved Tensions', *Southern Journal of Philosophy* **44**: 273–95. doi:10.1111/j.2041-6962.2006.tb00102.x.

Lippke, Richard L (2015) 'Elaborating Negative Retributivism', *Philosophy and Public Issues* **5**: 57–71.

Moore, Michael S (1997) *Placing Blame: A Theory of the Criminal Law*. Oxford University Press.

Moore, Michael S (2016) 'Responses and Appreciations', in Kimberly Kessler Ferzan and Stephen J Morse, eds., *Legal, Moral, and Metaphysical Truths: The Philosophy of Michael S. Moore*: 343–425. Oxford University Press.

Nielsen, Kai (1972) 'Against Moral Conservatism', *Ethics* **82**: 219–31.

Quinton, A M (1954) 'On Punishment', *Analysis* **14**: 133–42.

Rawls, John (1955) 'Two Concepts of Rules', *Philosophical Review* **64**: 3–32.

Smilansky, S (1990) 'Utilitarianism and the "Punishment" of the Innocent: The General Problem', *Analysis* **50**: 256–61. doi:10.1093/analys/50.4.256.

Wellman, Christopher Heath (2017) *Rights Forfeiture and Punishment*. Oxford University Press.