Law and Philosophy (2024) 43: 187–213 © The Author(s). This article is an open access publication 2023
https://doi.org/10.1007/s10982-023-09484-y

RICHARD CHILD

THE UNILATERAL AUTHORITY THEORY OF PUNISHMENT

(Accepted 24 May 2023)

ABSTRACT. It is frequently argued that wrongdoers forfeit, through their wrongdoing, their previously held claim rights against being punished. But this is a mistake. Wrongdoers do not forfeit their claim rights against being punished when they violate rights. They forfeit their *immunity* to having their claim rights against being punished removed. The reason for this, I argue, is that when they violate rights, wrongdoers culpably disregard the authority of right-holders to negotiate the conditions under which it is permissible to interact with them. The effect of this, far from undermining the authority of right-holders, is to transfer authority to right-holders to unilaterally impose the 'conditions of interaction' on wrongdoers *after* the violation. The conditions can be imposed for a diverse range of reasons and can take a variety of forms, including punishment. In this essay I explain and defend this new 'unilateral authority theory' of punishment.

I. INTRODUCTION

To show that punishment is morally permissible it is not enough merely to show that something valuable – crime reduction, retribution, reform of the offender, restitution for the victim, etc. – will result from the punishment. It must be shown that the person who is punished lacks a right against being treated in that way. Non-consequentialists tend to approach this latter task by trying to show that the act of violating another's rights triggers a significant change in the normative situation of the wrongdoer such that they ultimately

lose their previously held right against being punished.[1] I believe that this is, broadly speaking, the correct argumentative strategy for establishing the permissibility of punishment. But existing versions of this strategy all suffer from the same flaw: they mischaracterise the precise nature of the change that occurs in the wrongdoer's normative situation as a result of their wrongdoing.

In this essay I propose a new version of this general non-consequentialist strategy for justifying punishment that avoids the flaw in existing accounts. On my view, a wrongdoer does not automatically lose his right against being punished the moment he violates someone else's rights. Instead, he loses his right against being punished when that right is removed via the exercise of a normative power that is vested in the person whose rights have just been violated (or her representative).[2] The normative power that the victim (or her representative) has over the wrongdoer is the power to determine the costs the wrongdoer must bear for interacting with her in a way that violates her rights. And this power, in turn, is derived from the power we all have as right-holders to negotiate with others over what we can call the 'conditions of interaction', that is, the conditions that others must satisfy in order to gain permission to interact with us in ways that would, were permission not granted, violate our rights. By culpably violating another's rights, a wrongdoer disregards the authority of the right-holder to negotiate the conditions of interaction. But far from undermining the authority of the right-holder, the wrongdoer's disregard for that authority simply frees the right-holder from the reciprocal requirement to respect the *wrongdoer's* authority. Negotiation over the conditions of interaction is no longer required, for the victim of the violation can now assume

[1] The most obvious examples of such theories are so-called 'rights forfeiture' theories of punishment (see, e.g., Christopher Heath Wellman, *Rights Forfeiture and Punishment*, (Oxford: Oxford University Press, 2017) and Stephen Kershnar, 'The Structure of Rights Forfeiture in the Context of Culpable Wrongdoing', *Philosophia*, 29, (2002): 57–88). But there are many others, including: some versions of retributivism (e.g. Michael Moore, *Placing Blame*, (Oxford: Oxford University Press, 1997); consent theory (e.g. C.S. Nino, 'A Consensual Theory of Punishment', *Philosophy and Public Affairs*, 12 (4), (1983): 289–306); threat-based theory (e.g. Warren Quinn, 'The Right to Threaten and the Right to Punish', *Philosophy and Public Affairs*, 14 (4), (1985): 327–373); trust-based theory (e.g. David Hoekema, 'Trust and Obey: Toward a New Theory of Punishment', *Israel Law Review*, 25 (3–4), (1991): 332–350); and fairness theory (e.g. Herbert Morris, 'Persons and Punishment', *The Monist*, 52 (4), (1968): 475–501, and George Sher, *Desert*, (Princeton: Princeton University Press, 1987). Two further approaches that also belong on this list – Daniel McDermott's moral debt theory and Victor Tadros's duty view – are the focus of the critical part of my argument below.

[2] The significance of this 'or her representative' qualification will become apparent later in the essay. For ease of exposition I mostly omit the qualification from the early sections of the text.

unilateral authority over setting and imposing these conditions. The costs the victim may impose on the wrongdoer can be imposed for a diverse range of reasons and can take a variety of forms, including (but not limited to) punishment.

This is a very brief summary of what I call the 'unilateral authority' theory of punishment. In what follows I explain the basic ideas underlying the theory in more detail (Section 4) and respond to some important objections (Section 5). Before that, however, I critically analyse two of the most sophisticated competitor theories from the existing literature and show where they go wrong (Sections 2 and 3). Both of the theories I analyse (Daniel McDermott's 'moral debt theory' of punishment and Victor Tadros's 'duty view') agree with me that the key to justifying punishment lies in working out what we owe to people in virtue of their status as right-holders.[3] Where McDermott and Tadros go wrong, however, is in identifying specific goods that wrongdoers must now provide, or specific behaviours they must perform, in order to respond appropriately to their failure to give right-holders what they are owed. As I will go on to argue, these 'first order' responses to wrongdoing miss the crucial point that what right holders are owed in general is the 'second order' value of respect for their authority. The most appropriate way to respond to a failure to provide someone with this particular second order good is to let them decide, within certain limits, what the appropriate response is.[4]

The unilateral authority theory agrees with other non-consequentialist approaches that a significant normative change occurs at the moment of a rights violation. But rather than hold that this normative change directly affects the wrongdoer's claim rights

[3] Daniel McDermott, 'The Permissibility of Punishment', *Law and Philosophy*, 20 (4), (2001): 403–432. Victor Tadros, *The Ends of Harm*, (Oxford: Oxford University Press, 2011).

[4] The point I make in the second half of this paragraph echoes a similar point made by Malcolm Thorburn in his excellent discussion of the link between criminal punishment and the state's right to rule: 'What, then, should be the appropriate remedy to vindicate the state's exclusive right to rule? We are not concerned with the particular goods we may bring about through a particular remedy. Rather, our concern is what sort of remedy is required by the very idea of a state's exclusive right to rule'. Malcolm Thorburn, 'Criminal Punishment and the Right to Rule', *University of Toronto Law Journal*, 70, Supplement 1, (2020): 44–63. I address how my argument relates to Thorburn's work towards the end of Section 1.

against being used or interfered with, the unilateral authority view holds that the change affects the wrongdoer's *immunity* against any changes to his claim rights against being used or interfered with. The unilateral authority view is therefore an example of an 'immunity forfeiture' theory of punishment.[5]

This 'immunity forfeiture' approach is not only more theoretically satisfying than existing 'claim-right forfeiture' views, it also does a better job of explaining practice. One well-known problem with claim-right forfeiture views is that they have trouble ruling out vigilantism. The unilateral authority theory has no such trouble because it denies that immediately after a crime has been committed the offender automatically loses their claim rights against being punished. Instead, they lose their immunity against the removal of their claim rights against being punished. And this makes it much easier to explain why, in practice, only the state may punish. Upon committing a crime, the offender gains a moral liability that correlates with a moral power held by the state. The state exercises this moral power when it finds the offender guilty and sentences him, thereby further altering the offender's normative situation *in very specific ways* (i.e. taking away liberties, introducing new duties to do what the sentence requires, and granting new liberties to specific officials to carry out the sentence). Of course, the state's moral power is exercised via legal mechanisms in the name of the law. Nevertheless, when it comes to the question of whether state punishment is *morally* justified, it is the moral underpinning of these legal mechanisms that is crucial. The unilateral authority theory offers a much more accurate and compelling account of these moral underpinnings than existing rights forfeiture theories.

Before I begin the argument, one final preliminary point on methodology. The approach to justifying state punishment that I

[5] An immunity forfeiture view is one that holds that wrongdoers, at the moment at which they violate another's rights, forfeit their immunity from having changes made to their claim rights against being interfered with or used in various ways. The unilateral authority theory is just one of many possible theories explaining and justifying the idea of immunity forfeiture. The need for such explanation and justification seems obvious, but it is not always provided by proponents of the alternative claim-rights forfeiture view. For this criticism of claim-rights forfeiture views see Massimo Renzo, 'Rights Forfeiture and Liability to Harm', *The Journal of Political Philosophy*, 25 (3), (2017): 324–342.

adopt here – which begins with an account of how punishment is justified in interpersonal contexts before applying the analysis to the state – has been criticised recently in a series of articles by Malcolm Thorburn. Thorburn rejects the 'legal moralist' approach, as he calls it, in favour of his own more political approach. The main mistake made by legal moralist approaches, according to Thorburn, is that they attempt to answer the question, 'what purpose does criminal punishment serve?', when instead the question that theorists of punishment should be trying to answer is the conceptually prior question, 'what is required by the state's claim of practical authority over its subjects in the first place?'[6] Thorburn's own answer to this latter question is that the state can only claim practical authority over its subjects when it is able to 'vindicate its right to rule in the face of attempts to usurp it'.[7] Criminal punishment is justified, therefore, as the conceptually necessary 'legal mechanism' through which the state re-establishes its authority when its directives have been intentionally disregarded.

Thorburn's work is innovative and illuminating, and his emphasis on the role that punishment plays in vindicating authority clearly has significant overlap with the unilateral authority theory. But the very fact that there is this overlap suggests that Thorburn is too quick to claim that theorists of punishment must focus exclusively on the question of *state* authority. Indeed, one of the aims of this article is to respond to Thorburn's criticisms of legal moralist approaches by showing that we can justify punishment as a necessary implication of the concept of authority without appealing to specifically *state* authority. In order to do this I will assume throughout the discussion that we can talk about rights and right-holding in the abstract and independently of any state apparatus. Critics such as Thorburn, who invokes Hobbes and Kant in his work and who clearly believes that rights depend on state authority, may well object to this assumption. Unfortunately, to defend it here would take us too far off track. But towards the end of the article I will suggest how the argument I set out below generalises to include not only state of nature scenarios

[6] Thorburn, 'Right to Rule', 45–46.

[7] Thorburn, 'Right to Rule', 60.

but also civil society scenarios. If the argument is successful then it will go some way to reconciling legal moralists and their critics.

II. THE MORAL DEBT THEORY

In order to lay the groundwork for a more in-depth discussion of the unilateral authority theory, it is helpful to see where and how existing theories go wrong. The two theories I critically analyse in this section and the next section both agree that the right way to justify punishment is to focus on the respect that is due to persons in virtue of their status as right holders. By seeing how they misinterpret what exactly such respect requires we will be in a better position to understand the advantages of the unilateral authority theory.

According to McDermott's moral debt theory, justified punishment involves the forced removal of a special category of 'moral' goods that the offender is no longer entitled to in virtue of their prior wrongdoing. McDermott begins his defence of this claim by noting that when you violate my right to X you incur *two* distinct debts towards me. The first debt corresponds to the harm I suffered because of the violation. This debt can often be cashed out, literally, in terms of financial compensation. For example, if you stole $1,000 from me, you now owe me $1,000 (plus some extra, possibly, for any inconvenience and psychological distress caused). The second debt corresponds to the value of the treatment that you owed me but did not grant me in virtue of my status as a right-holder.[8] My right to my $1,000 was a right not only to the use of the money but also to a certain kind of respect from other people, as the legitimate owner of that $1,000. You therefore owe me a further debt for failing to provide this respectful treatment. And crucially, this further debt is *not* payable financially. Indeed, McDermott argues that this moral debt is *only* payable via punishment. Why is this?

McDermott's answer is that money is a 'material good' whereas being treated as a right-holder is a 'moral good'. Money is therefore 'not valuable in the *same way* as the treatment the wrongdoer

[8] McDermott, 'Permissibility of Punishment', 411.

withheld from his victim'.[9] What is special about being treated as a right-holder is that it distinguishes the recipients as persons, that is, as 'members of our moral community'. Attempting to settle a moral debt by transferring money therefore fails because 'providing others with material goods does *nothing* to distinguish them as members of our moral community'.[10]

So, ideally wrongdoers should transfer moral goods to their victims to pay their moral debts. But unfortunately, says McDermott, this is not possible. Moral goods, such as liberties and opportunities, friendship and love, are non-transferrable, and so there is nothing we can do to restore the victim by transferring moral goods. However, the fact that moral debts can't be paid *to victims* of wrongdoing does not mean that moral debts are normatively redundant. That you are unable to transfer to me the moral goods you owe me does not entail that you thereby maintain your entitlement to those goods. Indeed, the opposite is true. You have forfeited your entitlement to whatever goods would constitute payment of the moral debt irrespective of your inability to transfer them. And it's this fact, ultimately, that renders you liable to punitive treatment: 'Punishment, according this view, is a means of denying these forfeited moral goods to the wrongdoers'.[11]

There is much to be said for McDermott's theory. It offers a strong rebuttal to the pure restitutionist position which says that compensation exhausts the permissible responses to wrongdoing. And unlike many theories of punishment that focus on how the normative situation of wrongdoers changes when they violate rights, the moral debt theory has a very clear story to tell about how wrongdoing leads directly to the loss of the wrongdoer's claim rights against being punished. But in my view the most significant contribution of McDermott's theory lies in its focus on the special normative status of *being a right-holder*. In McDermott's view, what is special about being a right-holder is that you are owed a particular kind of moral good from others and that when these others fail to accord you this good they become indebted to you in a very specific, punishment-justifying way. I believe this thought gets close to the truth of the matter about the grounds of justified punishment, but it

[9] McDermott, 'Permissibility of Punishment', 414, emphasis in original.

[10] McDermott, 'Permissibility of Punishment', 418, emphasis in original.

[11] McDermott, 'Permissibility of Punishment', 424.

faces a serious problem which means that ultimately it must be rejected.

The problem is this: even if McDermott is right, and money is in general a material rather than moral good, the payment of money to a victim can still settle a moral debt if the relevant *authorisation* is given.[12] And in fact McDermott himself concedes as much: 'to say that a wrongdoer cannot *unilaterally* settle the debt by transferring money or property to the victim is not to say that the victim could not agree to accept money or property in lieu of punishing the wrongdoer'.[13] But this concession undermines McDermott's whole account because it demonstrates that what is really doing the work in his argument is not the idea that wrongdoers forfeit their claim to a rather mysterious set of 'moral goods', but instead the idea that wrongdoers render themselves liable to the normative authority of those whose rights they have violated.

Further support for this claim comes from the same footnote from which the quote in the previous paragraph was taken.

[12] I say 'even if McDermott is right' about money being a material good, because it's not actually clear why you can't repay (many, if not most) moral debts with money after all. Consider what McDermott says about what is distinctive about imprisonment: 'When we imprison a wrongdoer, we deprive him of a range of liberties and opportunities, many of which we only provide to other members of our moral community…we [no longer] allow them to pursue a career among us, to choose where to live in our communities, to make contracts, to patronize businesses, participate in politics, to drive a car, to get married, to make friends, to go bowling.' ('Permissibility of Punishment', 428). McDermott's point here is that depriving someone of moral goods like these is only possible by physically removing them from society. And it is true that imprisonment imposes some distinct deprivations. Even those who are completely destitute can still technically make friends, walk into shops, and hold a driver's licence. So forcing a wrongdoer to pay compensation, rather than go to prison, is not going to deprive them of *exactly* the same range of liberties and opportunities. But every single option on McDermott's list is made much less eligible for the person with no money. From participating in politics to going bowling, money is the key that opens these things up as genuine possibilities. And the point here is not simply that money increases your ability to do things that you are technically free to do already. As G. A. Cohen has argued, in a capitalist society money *is* freedom – the more money you have, the more freedom you have, (G. A. Cohen, 'Freedom and Money', in Otsuka (ed.), *On the Currency of Egalitarian Justice*, (Princeton: Princeton University Press, 2011): 166–192). If that is correct (and for the sake of argument here I assume it is) the distinction McDermott makes between moral goods and material goods no longer holds. When you compensate me for an earlier rights violation you don't just boost my material prospects, you also provide me with more of the freedoms which McDermott claims are distinctive moral goods. 'One of the most important ways in which we treat others as persons', says McDermott, 'is by allowing them to move freely among us as 'one of us'.' ('Permissibility of Punishment', 428). This is the role that money plays, for better or worse, in contemporary society. And this shows that it *is* possible to pay moral debts with money. (For an alternative version of this objection to McDermott – one that argues that there are various easily available (though non-pecuniary) means of paying moral debts – see David Boonin, *The Problem of Punishment*, (Cambridge: Cambridge University Press, 2008: 150–151).

[13] McDermott, 'Permissibility of Punishment', 413, emphasis in original.

McDermott says, 'in all the cases I consider, I will assume that *the victims prefer* that the wrongdoers be punished in response to their crimes'.[14] Again, this shows that the real work in justifying punishment is being done by the appeal to the authority of victims. If the victims prefer compensation to punishment, that's what should happen. If they prefer that nothing is done to the wrongdoer, that's what should happen.

Should we therefore conclude that there simply is no distinctive 'moral good' owed to victims? No. There *is* a distinctive moral good owed to victims, but it is not McDermott's rather vague notion of being treated as a 'member of the moral community'. Instead, it is the much more specific notion of being treated as the person with the authority to negotiate their conditions of interaction with others. What victims are doing when deciding whether to insist on punishment, or payment of compensation, or nothing at all, is exercising their unilateral authority to determine the costs to wrongdoers of violating their rights. Thus it doesn't matter what *type* of good wrongdoers are forcibly deprived of. In fact it doesn't matter if wrongdoers are not deprived of anything at all. What matters is that responses to wrongdoing are under the control of the relevant authority (which in the first instance is the victim of the rights violation). Victims' exercise of a normative power is what imbues the response with the requisite moral significance.

Another way to explain this point is to ask why a wrongdoer who violates another's rights can't rectify the situation by unilaterally bestowing generous amounts of compensation on the victim.[15] The answer is that such unilateral attempts to compensate don't engage the will of the victim, and so, although they might improve the *material* situation of the victim, they do nothing to address the *normative* impact of the original offence. However, if an offer of compensation is authorised by the victim (i.e. accepted as adequate compensation under appropriate conditions), then it does address the normative impact of the original offence. Material goods can *become* moral goods via authorisation.

[14] McDermott, 'Permissibility of Punishment', 413, emphasis added.

[15] Victor Tadros briefly discusses this question in his response to Liat Levanon's critical commentary on Tadros's book, *The Ends of Harm*. Victor Tadros, 'Replies', *Jerusalem Review of Legal Studies*, 5 (2012): 99. And Nozick famously raises the same question in *Anarchy, State, and Utopia* (Oxford: Basic Books, 1974): 59. A point in favour of the unilateral authority theory is that it provides (what I take to be) a convincing answer to this question.

Clearly, we require a much more detailed account of what is meant here by 'authorisation'. And we also need to know how victims come to have such authority over those who wrong them. This is the task of Section 4 where I set out the unilateral authority theory of punishment in more detail and show how it builds on McDermott's ideas. Before that, I consider a different attempt to ground punishment on a conception of what is owed to people in virtue of their status as right-holders.

III. THE REMEDIAL DUTY THEORY

Victor Tadros shares McDermott's belief that wrongdoers owe a distinctly moral good to their victims – respect for their normative status as right-holders – and that punishment involves coercively removing or recouping from the wrongdoer *after* the rights-violation what was owed to the victim *before* the rights-violation. As I have already said, stated in general terms I think this is the right way to think about justifying punishment. But Tadros, like McDermott, has the wrong interpretation of what exactly is owed to the victim.

Tadros's argument begins with a claim about how we may permissibly harm those who wrongfully threaten us with harm. As Tadros puts it, when your wrongful action creates a threat of harm to me you 'must bear significant costs to prevent that threat from being realized'.[16] In this case your primary duty not to wrongfully create threats of harm to me has been violated and now you have a remedial duty to help protect me from the threat. What's more, I may enforce that remedial duty by manipulatively harming you.

What if the threat is realised and you succeed in wrongfully harming me? Does your remedial duty to protect me disappear? No, argues Tadros. All that happens is that the content of the duty changes because although you can no longer protect me from the threat you *can* provide other kinds of remedial support. One option is that you protect me from a future threat. If a third person, P, is threatening to harm me in a roughly similar way to the way you harmed me then you have an opportunity to rectify your earlier breach of duty by preventing P from carrying out his threat. If you do this then there is a sense in which you cancel out your original

[16] Tadros, *Ends*, 268.

wrongdoing because I will be no worse off than I would have been had you never existed. And Tadros adds that your remedial duty to protect me from future harms is enforceable – just as a I can enforce your remedial duty to protect me from threats that *you* create, so I can enforce your remedial duty to protect me from threats created by others.[17]

In this way Tadros begins to piece together a justification for punishment as a system of general deterrence. For one of the main ways to enforce a wrongdoer's remedial duty to protect their victim from future threats is to punish the wrongdoer in order to deter others from wronging that same victim in future. It is true that further argumentative steps are required to fully justify a system of general deterrence, because based on what has been said so far it is an open question whether we may punish wrongdoers in order to protect people *other than* their victims. But the details of these further steps needn't concern us here.[18] Instead I want to briefly note the benefits of Tadros's theory before raising an objection that has not been discussed in the extensive critical literature on the remedial duty view. Again, seeing where the theory goes wrong will be helpful when it comes to developing the alternative account of punishment I set out in Section 4.

As with McDermott's theory, there is much to be said in favour of Tadros's argument. It is a wonderful example of that ideal mode of philosophical argument whereby one takes a widely accepted premise – that wrongdoers are duty bound to protect others from threats they culpably create – and tries to show how, via plausible steps, we can arrive at a controversial conclusion – that punishment is justified on grounds of general deterrence. It also represents one of the most ingenious attempts to defend an instrumentalist conception of punishment entirely from within a non-consequentialist moral framework. But despite its ingenuity, I'll now argue that Tadros's

[17] Tadros, *Ends*, 277.

[18] To complete the justification for punishment as a system of general deterrence Tadros adds three further argumentative steps. First, wrongdoers have a duty to substitute for each other in order to protect the class of victims as a whole (*Ends*, 280). For example, if you and I have both wronged others, and neither of us can protect our own victim, we might still be able to discharge out remedial duty by protecting each other's victim. Generalising from this we can say that all members of the class of wrongdoers have a duty to submit to punishment in order to protect all members of the class of victims. Second, victims of wrongdoing can pass on their 'entitlement to protection' to people they care about, including those who have not been victims of any wrongdoing (*Ends*, 280–281). Third, all citizens, including victims of crime, have a duty to protect all other citizens as long as the costs of doing so are not unreasonable (*Ends*, 298).

account of the ends for which we can harm wrongdoers is ultimately flawed, and for that reason the theory should be rejected.

The objection focuses on the vagueness of the idea that sits at the heart of Tadros's account, namely, that what offenders owe to their victims is the duty to remedy the harm they caused.[19] Initially, Tadros suggests the remedy could take the form of monetary compensation. But he promptly rejects this idea because of its impracticality and instead introduces the notion of 'future protection'.[20] Wrongdoers can remedy the wrong by allowing themselves to be used in ways that will prevent the same level of harm befalling the victim in future. And as we've already seen, one way to institutionalise this form of remedy is to set up a system of punishment aimed at general deterrence.

I don't disagree with Tadros that providing future protection might be one way for wrongdoers to remedy a previous wrong. My objection is that, by insisting that this is the only non-compensatory way of remedying the wrong, Tadros places too heavy a restriction on the normative authority of victims. For suppose the victim of a violent assault doesn't want to use their attacker to deter others and instead wants to insist that he enrolls on a course of rehabilitation or engages with a process of restorative justice. Tadros's theory would rule out these options as impermissible uses of the wrongdoer even if these options are less harmful to the wrongdoer. But given that the imperative here is to remedy the wrong done to the victim, to insist that the remedy must take a form that the victim explicitly does not want it to seems not merely unnecessary but downright perverse. The wrong that the victim has already been subjected to is now compounded by a failure to heed their preferences for how to respond to that wrong.

It is no good responding to this objection by pointing out that victims have the normative power to waive their right to the fulfilment of the wrongdoer's remedial duties. For the problem raised by the objection is not that victims are unduly constrained by a

[19] The closest Tadros gets to defining the concept of remedy is when he says, 'if a person has a duty to *v* and he breaches that duty, he retains a duty to do the next best thing' (*Ends*, 276). But to say that an appropriate remedy is whatever 'the next best thing' is does little to clarify the concept.

[20] Tadros, *Ends*, 277.

requirement to demand performance of the wrongdoer's remedial duties.[21] The relevant problem is that victims are unduly constrained by only being granted the right to demand *future protection*. Instead, the objection holds, victims should also have the option to demand a much wider range of actions on the part of wrongdoers.

A better response is suggested by an argument made by Kimberley Brownlee. Brownlee argues that the duty to provide future protection which plays a foundational role in Tadros's theory can be fulfilled in a much wider variety of ways than Tadros himself recognises. Not only might these alternatives be less harmful to the wrongdoer than Tadros's preferred deterrence-focused punishments, it is possible that they won't involve harming the wrongdoer at all.[22] As long as it can be shown that the rehabilitation courses or restorative justice processes that the victim prefers *are* in fact security-enhancing, then she will be permitted, even on Tadros's theory, to pursue them. This partially alleviates the worry underlying the objection but it still doesn't get to the heart of matter because it simply denies what the objection affirms, namely, that victims can use wrongdoers in the pursuit of non-protective, non-security-enhancing ends.

A third response is to concede the force of the objection but deny that this undermines Tadros's conclusion in favour of deterrence-based punishment. One way to do this is to argue that the normative relationship between victims and wrongdoers is not all that matters. If it *was* all that mattered, then perhaps victims would be free to choose from a range of options, including non-security-enhancing options, when responding to wrongdoing. But, continues the response, victims have a standing duty to other people to pursue the security-enhancing option. This strikes me as a quite a plausible view to take. But clearly we have now moved a long way from Tadros's remedial duty theory. On the view we are now considering, the normative options available to the victim (at least before the introduction of the general duty to pursue security-enhancing options) extend far beyond merely the duty to remedy the original harm.

[21] As it happens, this particular response is not available to Tadros in any case because he thinks that victims *are* constrained by the requirement to demand performance of the wrongdoer's remedial duties.

[22] Kimberley Brownlee, 'What are the Duties in the Duty View?' *Jerusalem Review of Legal Studies*, 5, (2012): 67–69.

Notice that according to this modified version of Tadros's theory it no longer makes sense to say that the wrongdoer owes a remedial *duty* to the victim. Instead, what the wrongdoer has is a liability that correlates with the victim's normative power to place the wrongdoer under such a duty. Until this power is exercised the wrongdoer is not strictly under a duty at all to the victim. And it is no objection to this interpretation of the situation that the victim is also under an independent duty to impose a particular type of treatment on the wrongdoer. For this is merely a duty to exercise normative power in a particular way and for a particular end – precisely the kind of duty-constrained power that we see throughout existing institutions of criminal justice.

IV. THE UNILATERAL AUTHORITY THEORY

Let us take stock of the argument so far. The two theories discussed above both focus on what wrongdoers owe their victims in virtue of the status of victims as right-holders. What we have seen is that the ways in which these theories interpret what is owed by the wrongdoer to the victim are flawed. The flaw in McDermott's account is that it ignores the significance of victims *authorising* the response of wrongdoers to their wrongdoing. By ignoring the role of authorisation, and restricting his attention to cases in which victims prefer a particular type of response to wrongdoing, McDermott fails to notice that what is really doing the work in his argument is the value of vindicating the victim's authority as a right-holder, not the vague notion of 'treating them as members of the moral community'. The flaw in Tadros's account is that it unduly restricts the ends for which the victim may permissibly use the wrongdoer. If punishment is justified by reference to the idea of remedying the wrong done to the victim, then it is perverse to ignore the preferences of the victim when they don't fit within a narrow definition of what counts as a 'remedy'. It is true that there may be normative considerations that militate in favour of certain types of punishment (e.g. ones that enhance the security of third parties), but these considerations are independent of the question of what the wrongdoer owes the victim in virtue of the victim's status as a right-holder.

Here is the general conclusion we can draw from the preceding discussion. Rather than ask what wrongdoers *owe* to their victims,

we should ask what victims can *demand* of wrongdoers. On the face of it these may look like two sides of the same coin but, as we shall see, the shift of perspective has significant normative implications. It pushes us away from an attempt to identify a first-order good, even a very abstract one, that must be transferred from wrongdoer to victim, and instead encourages us to look at second-order questions of decision-making, control, and authority. At this most fundamental level of theorising about punishment we should leave space for victims to exercise a significant degree of choice and control over what counts as a suitable response to wrongdoing. The theory that I will now present is an attempt to satisfy this desideratum.

I begin with a brief discussion of the nature of rights. One way to think of a right is as a moral boundary between persons. When one wants to interact with other people or the things that they own one can either do so impermissibly, by crossing the boundary, or permissibly, by waiting until the boundary is removed. Take for example the moral boundaries represented by my ownership right over my house or my right to bodily integrity. One way you might respond to these rights is to wait for me to invite you into my house or offer to shake your hand. When I do this I remove the moral boundaries that would normally make it impermissible for you to do these things. A different way you might respond is to simply decide unilaterally to cross the boundaries by trespassing or assaulting me. The key difference between these cases is that when the interaction is permissible it is because I exercised a normative power to give you a liberty to do what you previously had a duty not to do, whereas in cases of unilateral boundary crossings this normative power remains unexercised because my will was not engaged and my consent not sought.

Unilateral boundary crossings come in different varieties. When you unilaterally cross my boundaries unintentionally or otherwise excusably, the fact that my will was not engaged and my consent not sought does not display any culpable disregard on your part for my normative authority. There are times when we can't reasonably avoid crossing others' moral boundaries, and although in such cases compensation might be appropriate, to demand some sort of response to the fact that the authority of right-holders was disregarded would be a mistake. However, when you unilaterally cross my

boundaries with some degree of intent and without an excuse a different response is required. In this case your failure to engage my will or seek my consent displays a more or less culpable disregard for my authority.[23]

What is the appropriate response to a wrongdoer's culpable disregard for the authority of the right-holder? As we have seen, the wrong answer is to start asking what specific type of first-order good or mode of behaviour the wrongdoer owes to the victim. Instead, we should appeal directly to the authority that the right-holder *still has* over the wrongdoer. As countless theorists of rights have pointed out, the fact that someone violates one of my rights does not mean I don't have that right. Likewise, the fact that someone disregards my authority over my rights does not mean I don't have that authority. But the difficulty here is this: what is there left for me to have authority *over* if my moral boundaries have already been unilaterally crossed?

One thing I no longer have the normative power to do is to render permissible *ex post* an impermissible act that has already happened. I can certainly forgive previous wrongs by withdrawing blame. But without the ability to time travel I can't change the normative status of an historical act any more than I can change its empirical outcome.[24]

What I *can* still do, however, is determine the cost to others of interacting with me. There is no temporal limitation on this aspect of my normative authority as a right-holder. Let me explain this idea in more detail.

Consider the situation in which you and I are contemplating an interaction having never interacted previously. In this situation it is within my prerogative to determine the conditions that you must meet before I agree to remove my moral boundaries with respect to

[23] I discuss the relevance for my argument of the fact that culpability comes in degrees in Section 5, below.

[24] If there is a temptation to think that I *can* change the normative status of historical acts, I believe this can be explained by the fact that the exercise of authority that permits boundary crossings is often either tacitly expressed or implicit in our actions or relations. Thus many of the boundary crossings for which explicit consent was lacking, but with which we see no problem, are not examples where authority has been exercised *ex post* but where the act was rendered permissible non-explicitly at some earlier point in time.

you. I might not insist on any conditions at all, of course, but, if I do want to impose conditions, then I have wide discretion regarding how demanding those conditions are.[25] For instance, if you want to take over ownership of my car, I might decide to gift it to you, offer a low price, or offer a much higher price. And if you want to sleep with me, I can agree without insisting on any conditions, insist on a few minimal conditions (e.g. spend some time getting to know each other), or insist on much more demanding conditions.

Clearly if I make the conditions for granting exceptions to my moral boundaries too demanding then you are unlikely to interact with me because it is within *your* prerogative to decide whether to meet the conditions and engage in the interaction. But all this changes when you decide unilaterally to cross my boundaries. When you take or do what you want without negotiating with me in good faith over what the conditions of the interaction will be then you hand unilateral authority over the determination of the conditions to me. That is to say that the normative authority that I have in virtue of my status as a right-holder does not simply disappear when you violate my rights. Instead you merely undermine your own ability both to negotiate over the conditions of interaction and to decline to satisfy those conditions.

There is, then, a sense in which my normative power is actually enhanced by your violating my rights, because my exercise of that power is no longer conditional on your agreement. Of course, I may still decide to exercise unilateral authority in such a way as not to impose costs on you. In that case your liberties and claim rights will remain unchanged as a result of your wrongdoing. Alternatively, I may decide to impose costs on you, thus altering your liberties and claim rights in various ways. But what *always* changes as a result of your culpable boundary crossing is that you forfeit your *immunity* against my altering your normative situation. This is why, as I said in the introduction above, we can think of the unilateral authority theory as a version of an 'immunity forfeiture' theory.

Faced with this account of how one person comes to have unilateral authority over another following a rights violation, some challenging questions immediately confront us. The most obvious of these concern the proportionality of the conditions and the type of

[25] Exactly how wide my discretion is, and whether this causes a problem for the unilateral authority theory, is an issue I address in Section 5.

conditions that I can impose on you. The theory is bound to be rejected if it implies that I can kill you just for pinching me, or that torture is an eligible response to wrongdoing. There are also more abstract questions about the theoretical underpinnings of the account. For instance, does it rely, controversially, on the will theory of rights? And given that the theory implies that victims can impose such a wide range of conditions on wrongdoers, in what sense does it count as a theory of *punishment* in particular? Finally, there are questions concerning what we might call the practicalities of the account. For instance, if I die before I get a chance to exercise my normative power over you (say, because you murdered me), does the wrongdoer get away with it? Under what circumstances can third parties exercise power over wrongdoers on my behalf? I say something in response to all of these questions below. I don't have space to address every question in detail – and of course there are many further objections that I do not have space to address at all – but I hope to say enough to establish that the theory has promise and is worth developing further.

V. OBJECTIONS

The most obvious objection to the unilateral authority theory is that it gives victims too much discretion over the costs that wrongdoers may be required to bear. I will address this objection in some depth. The objection holds that the unilateral authority theory sanctions responses to wrongdoing that are wildly out of proportion to the degree of harm caused. The problem arises because the theory grounds the permissibility of punishing wrongdoers not on the harm caused by the wrongdoer (a scalar variable) but on the failure of the wrongdoer to respect a particular aspect of the victim's status as a right holder (a binary variable). The particular aspect in question is the authority of right holders to negotiate the costs that others must bear before being granted permission to interact with them. We typically believe that right-holders have wide discretion over the costs that others might be asked to bear, and we tend not to worry about this because others can decline the interaction and therefore decline to bear the proposed costs. But, as we have seen, when one violates another's rights, one's ability to decline to bear the proposed costs is forfeited. One now becomes liable to the forced imposition

of costs, and the fact that there are no obvious limits on how demanding these costs might be starts to look concerning. For instance, my right to bodily integrity implies that, before you pinch me, I may legitimately insist that you meet some extremely demanding conditions in order to be able to pinch me permissibly. At that point you are still free to scoff at these demands and walk away. But if instead you go ahead and pinch me without meeting the conditions, the unilateral authority theory seems to imply that my right to bodily integrity entitles me to impose *ex post* costs on you that might be extremely demanding, not just financially but in terms of your freedom and possibly even your life. This is extremely counterintuitive.

One immediate response to this worry is that a lot depends on the logically prior question of how much discretion right holders have over the costs to others of *permissibly* interacting with them, prior to any violation. On the most permissive answer to this question, right holders have unlimited discretion, and this would generate the strongest version of the worry we're considering. But less permissive views are available. For example, some hold that right-holders must offer a 'just price' for certain goods, or at least a price that is non-exploitative.[26] Proponents of such views do not usually apply them across the board to all possible rights that a person might be thought to hold (there is no 'just price' for the permission to sleep with someone, for example). But for rights that fall within their scope – typically property rights – it would appear that the authority of right-holders to impose costs on wrongdoers is quite heavily constrained, and this would be a welcome result.[27]

A second response to the 'too much discretion' objection is that although it might *seem* as though the failure of the wrongdoer to respect the victim's status as a right holder is a binary variable – they either respect it or they don't – the reality is more complex. This is because the degree of a wrongdoer's liability to the normative power

[26] The theory of the just price is often linked to Thomas Aquinas although its roots can be traced even further back to Aristotle. The theory is not currently in vogue of course but more recent proponents have included Emile Durkeim, *Professional Ethics and Civic Morals*, (London: Routledge and Keegan Paul, 1957), and James C. Scott, *The Moral Economy of the Peasant*, (New Haven, Connecticut: Yale University Press, 1976).

[27] Indeed, if the unilateral authority is correct, it would provide a reason *in support* of just price theory, since the widely shared belief that punishment should not be disproportionate would give us a reason to believe that there are limits on the discretion that right holders have when it comes to setting prices.

of their victim depends on the wrongdoer's *culpability*, and culpability is not a binary concept. I do not have space to flesh out exactly how differing degrees of culpability might affect the wrongdoer's liability, but it stands to reason that if non-culpable right violators are not liable at all, and fully culpable right violators are fully liable, then there will be numerous wrongdoers in between these two extremes – e.g. those who, rather than being malicious, are 'merely' heartless, reckless, or negligent – for whom liability is reduced accordingly.[28]

A third response is that there may be independent moral principles we can appeal to that limit the discretion of right-holders. For instance, we could appeal directly to a principle of proportionality which would specify strict limits on the severity of victims' response to rights violations of different kinds. The problem with this move, however, is that it is objectionably ad hoc. It goes against the whole thrust of the unilateral authority account's insistence that the authority of victims over wrongdoers is grounded in the pre-violation discretion they have over the costs that people in general may be asked to bear in order to interact permissibly with them. Instead of appealing to an ad hoc principle of proportionality, then, we might appeal to a more general humanitarian principle to treat everyone, including wrongdoers, with respect or humanity.[29] We could then distinguish between a narrow view of the unilateral authority theory and a wide view. On the narrow view, there are very few limits on how victims may exercise their authority over wrongdoers. On the wide view, however, further considerations such as the general humanitarian principle become relevant, and the formerly extensive discretion of victims is constrained by the requirement to respect wrongdoers by avoiding imposing extremely disproportionate punishments. Exactly how such a supplementary

[28] For a conceptually rich discussion of the different types and degrees of culpable states of mind, see Matthew Kramer, *The Ethics of Capital Punishment: A Philosophical Investigation of Evil and its Consequences*, (Oxford: Oxford University Press, 2011): 188–203.

[29] Appealing to such a principle is also one way of responding to the distinct worry, which I don't discuss in the main text, that the unilateral authority theory licenses cruel and unusual punishments. A humanitarian principle might prohibit the infliction of such punishments even if they can't be ruled out by appeal to the unilateral authority theory on its own. For a nice discussion of how humanitarian claims limit the harms we may impose on others who are otherwise liable to those harms, see Joanna Mary Firth and Jonathan Quong, 'Necessity, Moral Liability, and Defensive Harm', *Law and Philosophy*, 31, (2012): 693–700. An alternative to this appeal to humanitarian concerns might be an appeal to a principle of 'civilisation' as proposed by Jeffrey Reiman, 'Justice, Civilization, and the Death Penalty: Answering van den Haag', *Philosophy & Public Affairs*, 14 (2), (1985): 115–148.

principle would work, and precisely what limits it would entail, is beyond the scope of the current essay.

A fourth and final response to the 'too much discretion' objection focuses on the comparison between the state of nature and civil society. In the state of nature, continues the response, the concern that victims would have too much discretion is well-founded. But this doesn't show that the unilateral authority theory is flawed; it simply highlights the drawbacks of the state of nature and the importance of moving from the state of nature into civil society. Once that move has been made, and the rule of law set up, victims can no longer exercise authority over wrongdoers in a direct and individualistic way. Instead, their authority can only be exercised indirectly and collectively via state institutions. As part of this shift from direct to indirect authority, at least in any state that respects the principles of the rule of law, the ability of victims to exercise authority in unpredictable and inconsistent ways will be undercut. Instead, the rule of law-respecting state is bound to put together a schedule of punishments that is stable, consistent, and publicised, as well as being ordinally and cardinally proportionate in a way that fits with the common sense of the citizenry.[30]

But does this not leave us without any critical purchase on the penal practices of particular societies? And does it not imply that different societies with completely different types and degrees of punishment for the same crimes, and indeed different reasons for wanting to punish, will all be acting equally within their rights? I see this as a feature of the view rather than a bug. Taken on its own, independently of any supplementary moral principles, the unilateral authority theory sees a society of abolitionists and a society of draconians as equally legitimate. The crucial caveat, of course, is that the unilateral authority theory is not exhaustive of morality. We have already discussed the possible supplementary role played by a humanitarian principle. Another relevant principle is the principle of non-discrimination. And a further relevant principle is the principle of democracy. The existing U.S. penal system falls foul of all three of these principles and any complete theory of political morality would therefore judge it to be illegitimate and unjust. But if the cruel and

[30] On the distinction between ordinal and cardinal proportionality, and how the two types of proportionality can be combined, see Andrew von Hirsch and Andrew Ashworth, *Proportionate Sentencing: Exploring the Principles*, (Oxford: Oxford University Press, 2005), section 9.3.

discriminatory nature of punishment in the U.S. was corrected, and the system made much more democratically responsive to the will of the citizens who live with it rather than the corporations and power brokers who stand to gain from it, then the fact that it is a relatively harsh system would not be nearly so objectionable as it currently is.[31]

This discussion of the state has served to connect theory to reality, however briefly. But I now want to return to pure theory and ask whether the unilateral authority theory presupposes the correctness of the so-called will theory of rights. One might think so because the will theory asserts that the person who holds a given right is the person who is 'empowered to make a choice about the fulfilment of someone else's duty'.[32] This sounds very close to the idea at the heart of the unilateral authority theory. And this could be a problem simply because the will theory is far from universally accepted. To make the unilateral authority theory contingent on the will theory would not prove that it fails (because the will theory might turn out to be correct), but it would place the theory on shaky foundations and undermine its attractiveness to proponents of the competing interest theory of rights (which asserts that a right-holder is someone who stands to benefit from the fulfilment of another's duty).

Fortunately, it is relatively easy to show that the unilateral authority theory does not depend on the will theory. For the unilateral authority theory does not presuppose that the only agents who can hold rights are those who have the authority to determine the conditions for permissible interactions. Indeed, the unilateral

[31] It would still be objectionable to many people, of course, purely on grounds of its harshness. But at this point, I believe we are into the realm of reasonable disagreement. There may be an objective truth of the matter regarding how best to treat wrongdoers but, unless it can proved what this objective truth is, our theory of punishment needs to accommodate good faith disagreement. This is one of the advantages of the unilateral authority theory. As a fundamentally procedural theory, it tells us what makes punishment legitimate, and therefore permissible, even if the specific punishments it sanctions are not, according to the correct objective standard, perfectly just. There are many difficult questions here, of course. One important issue for future research concerns the significance of the fact that, whereas in the state of nature imposing costs on a wrongdoer is something that an individual takes on either by themselves or with the voluntary help of others, in civil society imposing costs is something that everyone is forced to contribute to via taxation. I believe this fact may give us grounds for limiting the ways in which states can use wrongdoers to those that can be justified in terms of public reason, but I can't develop the argument here. For an early attempt to construct a contractualist theory of punishment based on public reason considerations, see Corey Brettschneider, 'The Rights of the Guilty: Punishment and Political Legitimacy', *Political Theory*, 35, (2007): 175–199.

[32] Matthew Kramer, Nigel Simmonds, and Hillel Steiner, *A Debate Over Rights*, (Oxford: Oxford University Press, 1998): 2.

authority theory is agnostic on the question of what qualifies someone as a right-holder. It claims only that the authority to punish ultimately derives from the authority to determine the conditions for permissible interaction. This is perfectly compatible with there being right-holders who lack authority in either of these two senses, and thus the unilateral authority theory is perfectly compatible with the existence of interest theory rights. This is important if the theory is to appeal to those who believe that, for example, children and people with severe disabilities have rights.

But isn't the existence of right-holders who lack the authority to render wrongdoers liable to punishment a serious problem for the unilateral authority theory? Consider children for example. When someone violates a child's rights it would be implausible to suggest that the child himself has authority to render the wrongdoer liable to punishment. Are we therefore forced to conclude that violations of children's rights must go unpunished? No, because sometimes the right-holder and the person with authority to determine the conditions for permissible interaction with the right-holder are different people. In the case of children this is simply the relationship between ward (right holder) and guardian (person with authority to determine the conditions for permissible interaction with the ward).

A similar story can be told about the relationship between citizens (right holders) and the state (agent with authority to determine the conditions of interaction).[33] Of course the analogy – between the child-parent relationship and the citizen-state relationship – is not perfect. One obvious difference is that children (especially young children) have very limited authority to determine the conditions of their permissible interactions with others, whereas adult citizens retain significant authority over the conditions of their interaction with fellow citizens. But here we need to make a distinction between *ex ante* authority and *ex post* authority. *Ex ante* authority is the authority a person has to determine the conditions of their interaction with another *before* any rights violations have occurred (e.g. by setting the price for goods or by controlling access to their bodies). *Ex post* authority is the authority a person has to determine their

[33] I acknowledge my debt here to Malcolm Thorburn's excellent analysis of the structural similarity between parental authority over children and state authority over citizens. See Thorburn, 'Punishment and Public Authority', in Antje du Bois-Pedain, Magnus Ulväng and Petter Asp (eds.), *Criminal Law and the Authority of the State*, (Oxford: Hart Publishing, 2017).

conditions of interaction with another *after* the other has violated their rights. In civil society (at least in liberal states) citizens retain a high degree of *ex ante* authority but give up most of their *ex post* authority.[34]

The idea of citizens 'giving up' authority to the state should not be interpreted too literally. Lockean consent theorists of state authority and political obligation might be comfortable with this notion, but others (e.g. Kantians) will object to the idea that citizens in civil society have actively given up authority that they would otherwise have had in the state of nature.[35] Once again, though, the unilateral authority theory is agnostic on these deeper questions of political philosophy. It is perhaps easier to see how the mechanics of the unilateral authority theory work on a Lockean view, with its highly individualistic approach to rights (which is why I have helped myself to this framework in setting out the account above). But the theory applies just as well on a Kantian view. The Kantian state might not be a fiduciary exercising authority on behalf of individual citizens, but it still represents their (collective) will. Its authority to determine the conditions of interaction between individuals still ultimately derives from the will of those individuals and so the unilateral authority theory still applies. Rights-violators in the Kantian state disregard citizens' collective authority and punishment is a way of reasserting that authority.[36]

This leads us onto the final objection I want to address. If wrongdoers only become liable to punishment through an exercise of authority, can wrongdoers avoid such liability simply by killing or otherwise incapacitating the victim? There are two parts to the answer to this question. First, in the state of nature, third parties may step in and exercise a victim's authority on their behalf as long as they do so in a way that makes a reasonable attempt to track the preferences of the victim. There is nothing mysterious about this. Whenever someone is incapacitated or deceased, another must assume authority to make decisions (e.g. medical or financial decisions) on their behalf, even when there is no explicit statement of will. Of

[34] In many legal systems citizens give up 'most' of their *ex post* authority rather than all of it because they are often still involved, to a greater or lesser extent, in both criminal and civil law proceedings.

[35] I'm grateful to an anonymous reviewer for this journal for pressing me on this point.

[36] I'm grateful to a different anonymous reviewer for this journal for suggesting a response along these lines.

course there are many complex issues here regarding the principles that guide the exercise of power of attorney that I do not have space to discuss.[37] But these internecine disagreements are consistent with, indeed dependent on, consensus on the general idea.

The second part of the answer to the question in the previous paragraph is implicit in what I said above about the move from the state of nature to civil society. One of the many advantages of civil society is that the state can represent us when we either can't or shouldn't represent ourselves. Precisely what justifies this claim is beyond the scope of the current discussion. But assuming there *is* a justification available then once again we can rely on rule of law constraints on the state's exercise of authority to solve any lingering problems that attend *ad hoc* attempts to represent the will of individuals in the state of nature.

VI. CONCLUSION

My aim in this article has been to show how punishment can be justified by appealing to the authority that right-holders have over the conditions that others must satisfy in order to interact permissibly with them. I first analysed two existing attempts to justify punishment, both of which share the unilateral authority theory's focus on the failure of wrongdoers to give victims what is owed to them as right-holders. I argued that both theories are undermined by overly narrow conceptions of what is owed by wrongdoers to victims. For what is owed to right-holders is not some first-order good that can be identified at the level of theory and then extracted from wrongdoers via some targeted institutional strategy. Instead, what is owed to right-holders is the second-order good of respect for their authority. And this is a good that wrongdoers can be forced to provide via a wide variety of different actions.

The rest of the article was devoted to elaborating this unilateral authority theory of punishment and defending it from various objections. I have only been able to address a handful of the many questions and complications that the theory confronts, but one thing that emerges very clearly from the discussion is how important the

[37] For a good discussion of some of these issues, focused on the role of the 'best interests' principle in medical contexts, see John Coggon, 'Best Interests, Public Interest, and the Power of the Medical Profession', *Health Care Analysis* 16, (2008): 219–232.

state is for resolving many of the problems of punishment in the state of nature. The discretion that the multitude of 'small-scale sovereigns' has in the state of nature is a source of great uncertainty, instability, and potential violence, and it generates a powerful reason to move out of that situation and into civil society.[38] Once we have made that move, a potentially wide range of supplementary principles comes into play to limit the ways in which we can exercise authority over wrongdoers. But the theory still allows for democratic authority to play a significant role in determining what counts as a just and legitimate punishment. This is a welcome result, and even if the details of the account I have set out are ultimately rejected, I hope to have bolstered the case for a greater focus on the concept of authority in the philosophical study of punishment.

FUNDING

No funding was received to assist with the preparation of this manuscript.

DECLARATIONS

COMPETING INTERESTS The author has no competing interests to declare that are relevant to the content of this article.

ETHICAL APPROVAL No ethical approval was required for this study.

OPEN ACCESS

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from

[38] HLA Hart, *Essays on Bentham*, (Oxford: Oxford University Press, 1982): 183.

the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

The University of Manchester, Manchester, UK
E-mail: richard.child@manchester.ac.uk

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.