SYMPOSIUM

THE PHILOSOPHY OF PUNISHMENT

A PRÉCIS OF *PUNISHMENT*

BY

THOM BROOKS

© 2015 – *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015): 3-23
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

A Précis of *Punishment*

Thom Brooks

Punishment is a topic of increasing importance for citizens and policy-makers. The same can be said for academic researchers and students. Mass imprisonment has reached record high levels while public confidence is often lacking. New thinking is required urgently to address these challenges. Moreover, there have been several key developments in the philosophy of punishment over the last 20 years absent in leading guides including the communicative theory of punishment, restorative justice and my novel unified theory of punishment.

My book *Punishment* is a critical introduction to the philosophy of punishment attempting to offer a new and refreshing approach to benefit scholars and students alike.[1] While the book is primarily philosophical, it brings together relevant insights from law, criminology, criminal justice, politics and sociology. The aim is to provide both a comprehensive overview with new insights on many familiar theories of punishment.

The book begins with a brief introduction clarifying what is meant by punishment and its relation to morality. The following first part of the book examines what I call 'general theories' of punishment. These are theories that have a single purpose or aim.

[1] Thom Brooks, *Punishment* (London: Routledge, 2012).

 Philosophy and Public Issues (New Series), Vol. 5, No. 1 (2015): 3-23
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

These include retributivism, deterrence, rehabilitation and restorative justice. The second part considers hybrid theories that attempt to bring together multiple penal purposes. The hybrid theories discussed are the mixed theory of Rawls and Hart, expressivism (including the communicative theory of punishment) and the unified theory of punishment. The final part of the book looks at how these different theories about punishment relate to certain case studies, such as capital punishment, juvenile offenders, domestic violence and sex crimes like rape and child sex offences.

In summary, I attempt to show why various theories of punishment attracts wide support and examine each in terms of theory and practice. I argue that each of the traditional theories of punishment has much to recommend it, but each also runs into real problems. My unified theory of punishment is my effort to show how we might bring together what is attractive about each of the other theories of punishment in a coherent framework, but without their problems.

This *Precis* will provide a brief overview of the book. The below sections cover the introduction and each of the three sections. My discussion is not exhaustive and only attempts to indicate to the general arguments and set the scene for considering the papers in this special issue that engage with my book.

The Introduction

I begin *Punishment* with an important definition. First, I argue that my discussion of punishment will be focused only on punishment *for breaking the law* (1). Punishment is a word used in

many different ways. Some speak of 'punishing' a misbehaving child or perhaps a pet. Or that a difficult physical activity like cycling steep hills can be 'punishing'. But these all point to different things. I am interested here only in the phenomena of punishment for a crime. Punishment by the state for a crime is different in form and content from these other activities. The parent who is said to punish a child does not do so because a law has been broken or even because the child has breached some rule he knew about in advance. Nor is there an appeal. Perhaps the only similarity between this idea of 'punishment' and (legal) punishment is both are impositions of some burden because of some earlier act or omission.

I believe this link between crime and punishment both crucial and too often overlooked: 'We will ask which theory of punishment is best if, and only if, a relevant law is justified. The possibility of justified laws reveals the horizon of just punishments'.[2] In other words, punishment presupposes a crime that is a trigger for the punishment. There can be no punishment without a crime—and the justification of punishment is bound up with that of its linked offence.

Punishment is *a response* to an offence. So when we think about punishment, we consider what should be the best response to an offence. This response must be of a person for breaking the law administered and imposed intentionally by an authority within a legal system that imposes a loss (4-5). Punishment would otherwise be arbitrary and, if not some form of loss, might become indistinguishable from rewards.

I further distinguish the definition of punishment from its aim and distribution, now a common feature of most analytic jurisprudence on this topic since H. L. A. Hart's *Punishment and*

[2] Brooks, *Punishment*, 3.

Responsibility.[3] Each asks different questions: how *is* punishment understood? What is the *aim* of punishment and how should it be *distributed* to individuals? My definition does not load the dice in favour or against any particular view. Retributivists, deterrent theorists and others can all accept the link between crime and punishment. But each will have different ideas about the purposeful aim of punishment, such as whether it should be deserved or deter. This will then impact on which individuals might be selected by a theory of punishment in order to fulfil its aims. Much of the book focuses on problems arising with the aims and distribution of various penal theories.

The Introduction closes with a consideration of two influential views about criminalisation that are relevant. The first is *legal moralism* and this is the idea that criminalization should be linked to immorality. The problem I raise is that legal moralism is undermined by what I call *the naturalist fallacy*: 'there is no necessary connection between crime and immorality, even if there is often this connection'.[4] My point is morality is no certain guide to identifying all crimes we would want linked to punishment. If no single view of morality can produce a list of all the crimes we would want to punish, then we must find some alternative. I am highly critical of moralistic and natural law-friendly views of the criminal law and sentencing throughout the book.

A second influential view is the *harm principle*. This is the principle first stated clearly by John Stuart Mill that the only purpose for which we may restrain someone is to prevent harm to others. This view often links harm with other-regarding harms, or harms that are imposed by one on another. Self-regarding harms are often missed. My criticism with this perspective is that

[3] H. L. A. Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (Oxford: Clarendon, 1968): 4.
[4] Brooks, *Punishment*, 10.

harm *simpliciter* is not compelling because not all harms are or should be criminalised, such as injuries suffered during the normal course of a contact sport or receiving surgery. My point is that morality and harm may have importance, but they do not determine in any obvious way what should be criminalised and, therefore, punished without qualification.[5]

Part 1 – General Theories

The first substantive part of the book considers four general theories about punishment. Each has something highly compelling at its core. Retribution gets right the importance of desert: that an individual must have done or omitted something to warrant punishment. Offenders must be deserving of their punishment. Punishment is not private vengeance, but public justice. Few would disagree with retributivists that the innocent should never be punished.

But this does not mean we would all agree on what is or is not 'deserved'. Retributivists often make serious mistakes in linking desert with some view of moral responsibility for wickedness. The greater the responsibility for an immoral wrong, the more a person 'deserves' punishment. But the problem is that not all crimes are linked to immorality in this way. Some crimes are strict liability offences where might be at best *causally* responsible, but moral responsibility is irrelevant. Not all crimes are evil and some might even be thought amoral. Even if we did think offenders should be punished to the degree they are morally responsible for some immoral deed, we cannot read the minds of others. This is

[5] See Thom Brooks, 'Criminal Harms' in Thom Brooks (ed.), *Law and Legal Theory* (Boston: Brill, 2014): 149—161.

important for retributivists because an individual's desert, at least on a classical 'positive' view of retributivism, should be determined entirely in accordance with an individual's mind-set at the time of an offence. Our best guesswork is not good enough.

There are also related issues about proportionality and desert. It must be noted that retributivism is a remarkably wide tent covering a diverse range of perspectives. Nothing brings this out more than considering its diverse views of 'retributivist' proportionality. For example, if someone should only be punished to the amount deserved (as some versions of retributivism claim), then this would suggest some form of strict equality between the crime and its punishment. However, this cannot be compelling for at least two reasons. First, this would render most crimes unpunishable. There is no like for like punishment available for many victimless offences like drug possession, speeding or perhaps even theft.[6] Secondly, even where we could do like for like, there are strong reasons against doing so. Capital punishment might be a controversial case, but sex crimes and torture are not: there is no reasonable advocate for doing unto others as they've done to their victims when it comes to these violent offences. *Punishment* considers these and other challenges for retributivism. My conclusion is that retributivism gets some things right like the importance of an offender's having committed an offence as central to whether or not that person is punished. But we must look elsewhere for a more plausible view of how crime and punishment might be linked up.

The next chapter considers deterrence. This is the view that punishment is justified by its deterring potential offenders in

[6] Consider theft. If I still your bicycle, then how might I be punished like for like if I do not have a bicycle or any comparably similar possession another might take from me?

future. This can be understood in terms of *macrodeterrence* where our aim is to create a deterrent effect among the general public or as *microdeterrence* where we aim to deter specific individuals. I also discuss this view with respect to incapacitation that crime can be reduced through imprisoning offenders. I argue that deterrence theorists broadly get right the importance of crime reduction: few of us would prefer a criminal justice system that made crime more likely.

But there remain significant questions about how this might work and I raise a number of problems. The first is the *problem of geography*: this is the false belief that crimes only happen outside prisons. Since crime can occur in prisons too, then putting offenders in prison does not mean they cannot perform crimes while they are incarcerated. A second problem is that wrongness does not play any fundamental role for deterrent theorists. Much as retributivism is perhaps burdened by its controversial moralistic commitments, deterrence is rendered problematic by its lack of any such commitment. In effect, deterrence is about telling us how much we should punish and not what we should punish. While retributivists can speak about what might be deserved, deterrent theorists are agnostic on the wrongness of crimes. We deter not in the degree an offence is wrong, but in terms of what might motivate others to avoid committing such an offence in future. This opens deterrence up to what I call the *problem of time and changing effects*: what might deter today may not deter tomorrow. So while other penal theories may seek a more fixed view of crime and punishment, deterrence can recommend a different punishment for the same crime as what would be required to deter changes over time. This is further complicated by the *problem of difference*, namely, that different people may react very differently to the same deterrent effects. But the biggest problem of all is whether we can know deterrence works. We can measure how many crimes were recorded, but can we ever know

how many crimes *might* have happened if punished did not have some deterrent effect? I illustrate this in class by asking students how many did not steal a bicycle on the way to the lecture *because they feared punishment*. In about every case, students say they avoid theft not out of the fear of punishment but because they do not want to steal anyway. And so evidence of crime reduction, if proven, might still be no evidence that deterrence has worked.

Chapter 3 focuses on rehabilitation. This is the idea that the great majority of offenders will one day leave prison and so prison should be used to assist their transition from criminal to law abiding citizen. Rehabilitative punishments can take many different forms such as therapeutic treatments like cognitive behavioural therapy or recreational therapy, but also education and training.

I raise several problems for rehabilitative theories. One is the role of morality. Most rehabilitative theories view themselves as some form of moral education.[7] But rehabilitation does not target every moral wrong, only those that are criminalised—and not all crimes are clearly immoral. So understanding rehabilitative punishment as a form of moral education makes for a poor fit with the criminal law. Another problem is that individuals committing the same crime in similar circumstances might be punished very differently depending on how quickly they might be morally educated. The murderer who is deeply and sincerely repentant might then appear to demand less punishment on this view than an unapologetic pickpocket because the latter will take much longer to convince of his need to reject his criminality. This links up with the problem of the unreformable: those who are resistant to reform. It is implausible to think they should be punished most of all no matter how trivial their offence.

[7] See Brooks, *Punishment*, 56-57.

Part 1 closes with a chapter on restorative justice. Restorative justice is different from other approaches. Retribution, deterrence and rehabilitation theories of punishment are typically conceived within a traditional setting of sentence and offender. Restorative justice is an alternative to the formal courtroom setting and sentencing procedures that prioritises informality and dialogue. It typically takes the form of victim—offender mediation or a conference setting where a trained facilitator manages a conversation between victim, offender and others. The purpose is to bring about greater mutual understanding and to 'restore' the status of the offender from lawbreaker back to full citizen. Restorative justice requires offenders to have acknowledged their wrongdoing and make some apology to victims. The results are promising: studies have shown participants show high satisfaction, there is less reoffending and all at lower costs.

But this masks some problems. Not all victims want to take part—and likewise not all offenders. Restorative justice might not even be thought to be a theory of punishment. This is because it rejects the use of prison to bring about restoration. A consequence is that it is often reserved for minor offences committed by youths. So its restricted set of possible outcomes limits its applicability to more types of crimes and offenders. There are also serious questions about who is being 'restored' to who and even what is being 'restored'.

These comments are a broad overview. There are many other points made and positions considered with further objections all in much greater detail. But I only wish to provide some indication of a few main points raised. The discussion is meant to show that each view gets something right. Retributivists are right that desert matters even if there are problems with how desert is understood by them. Deterrent proponents are right that crime reduction is an important goal even if we might not ever know if it was

brought about by a threat of punishment. Rehabilitation gets right that most offenders will one day leave prison and it can be crucial to assist their transition to law abiding citizen, otherwise we risk rendering such individuals even worse off at our peril. Restorative justice gets right its effects of fostering equality and dialogue with impressive results, but has problems with who should take part and what is being restored.

Part 2 – Hybrid theories

This discussion leads us to next consider three different ideas about how these different purposes might be brought together into what I call hybrid theories of punishment.

The first I cover in a chapter 'Rawls, Hart and the mixed theory'. Rawls and Hart endorse different ideas about punishment, but share a similar core. This is the idea that the legislature looks forward in setting out what is criminalised and how much it might be punished. This forward-looking perspective is utilitarian in its outlook. In contrast, the judiciary is backward-looking and retributivist in considering what *this* person might deserve for some past action. The main claim is that punishment brings together both forward-looking and backward-looking perspectives. They do not clash because they are considered at different points: the one when we think what should be punished, the other when we punish a particular person.

A problem with this view is that if offenders should only be punished as much as is deserved and as distributed by judges, it is then difficult to see how there is to be a deterrent effect on the whole. If everyone gets what is deserved, then there might not be

any deterrent effect unless what is deserved *also* deters. But either way what counts is desert and not deterrence. So the two do not have an equal status and one has more importance than the other.

I discuss this problem in the context of negative retributivism. This is the idea that desert is necessary for punishment, but not sufficient: whether or not we punish should be determined by non-retributivist factors. While this view has its vocal proponents, it is also conceptually incoherent. If desert is so vitally important that only it should matter for choosing who might be selected for punishment, why should it *necessarily* be not crucial for determining punishment's amount?

A second hybrid theory considered is expressivism. This view has many proponents, but its leading modern defender is Joel Feinberg. He argues that punishment was different from penalty in kind. Punishment is said to be prison and penalties other forms of sanctions. For Feinberg, punishment as imprisonment requires something different in its justification from mere penalties. He claims this is punishment's expressivist effect: that it can express public denunciation for performing a wrong.

This expressivist model is developed further by Antony Duff. He argues in what he calls his communicative theory of punishment that punishment is not only about the public expressing its denunciation to offenders, but offenders communicating their apologies back. Punishment is not a one way street, but a dialogue. Both Feinberg's and Duff's models are thought to be consistent with desert, to provide a deterrence and motivate rehabilitation. Only those persons deserving of public denunciation are selected for punishment, this is a message that citizens will not wish to receive and so avoid criminality and to be subject to such a message can give reason for what Duff calls 'secular penance'. In these ways, expressivism aspires to be a hybrid theory.

I raise several problems with this perspective. The first is that the commonly drawn line between punishments and penalties is too sharp and fails to reflect how sentencing actually works. Offenders do not face an option of prison *or* some alternative, but often some combination. A prison sentence can include a reparation order, for example. So to say that expressivism is about justifying punishment exclusively addresses only one part of how sentencing works. Moreover, there is no convincing reason given as to why a fine or community sentence cannot be understood as an expression of public denunciation. Any state imposed sanction can be understood in this way.

A second problem relates to communicative theories in particular. They argue that offenders 'communicate' an apology back to the community by serving their sentence as a kind of secular penance—and this is true whether or not the offenders does, in fact, communicate anything at all either way. I argue in *Punishment* that:

> If it does not matter whether any offender repents and all repentance is at minimum assumed, then what is the clear difference between retributivists and communicative theorists? Is the difference little more than that the latter *assume* that offenders repent through serving time in prison? […] A theory that says it's justified because offenders repent and they repent because it's assumed by the theory is not compelling.[8]

Even worse, there is no evidence that repentance is best served through imprisonment any way. In any event, expressivists do not actually justify punishing offenders by as much as the public does, in fact, wish to express its denunciation of their acts. What counts most is what offenders deserve: if the public wished for a more punitive sentence to send a message, this could breach what is deserved and lack support from expressivism. So public

[8] Brooks, *Punishment*, 120.

censure may be an important aspect of punishment, but it is unclear if it can and should serve more than a metaphorical role even by expressivist standards.

In summary, the book has surveyed these two major attempts at bringing multiple penal purposes together and found them unsatisfactory. But is there a model we can look to instead? I believe there is and I call it *the unified theory of punishment*. The first thing to note is that multiple penal purposes are a regular feature of sentencing guidelines. This might be traced back to the influence of the 1962 Model Penal Code that claimed sentencing had several justificatory principles such as retribution, deterrence and rehabilitation. This is echoed elsewhere, such as in section 142 of the Criminal Justice Act 2003 in England and Wales. The problem is that the judges and magistrates who determine sentences lack a framework for weighing these different penal purposes in a coherent and unified way.

I argue that a unified theory of punishment is not only possible, but compelling. But I am not the first to try. Credit must be given to Hegel and the British Idealists, as I argue in *Punishment* and elsewhere.[9] What they got right was a coherent, unified

[9] See Brooks, *Punishment*, 126—127; Thom Brooks, 'T. H. Green's Theory of Punishment', *History of Political Thought* 24 (2003): 685—701; Thom Brooks, 'Is Hegel a Retributivist?' *Bulletin of the Hegel Society of Great Britain* 49/50 (2004): 113—126; Thom Brooks, 'Does Bevir's *Logic* Improve Our Understanding of Hegel's *Philosophy of Right*', *The European Legacy* 11 (2006): 765—774; Thom Brooks, 'Punishment and British Idealism' in Jesper Ryberg and J. Angelo Corlett (eds), *Punishment and Ethics: New Perspectives* (Basingstoke: Palgrave Macmillan, 2010): 16—32; Thom Brooks, 'Is Bradley a Retributivist?' *History of Political Thought* 32 (2011): 83—95; Thom Brooks, 'What Did the British Idealists Ever Do for Us?' in Thom Brooks (ed.), *New Waves in Ethics* (Basingstoke: Palgrave Macmillan, 2011): 28—47; Thom Brooks, 'Punishment: Political, Not Moral', *New Criminal Law Review* 14 (2011): 427—438; Thom Brooks, 'Hegel and the Unified Theory of Punishment' in Thom Brooks (ed.), *Hegel's Philosophy of Right* (Oxford: Blackwell, 2012): 103—123 and Thom

theory of punishment requires a new framework. Different penal purposes may clash and so there must be some way to manage potential conflict. For example, the aim of punishing offenders as much as they deserve might clash with the aim of deterring others: the amount of punishment deserved by one aim might be insufficient to satisfy the aim of the other. It is because this penal pluralism can lead to conflict that we require a new framework. For reasons I will not pursue here, I argue that the reasons offered by the Hegelians on how to provide this framework are unsatisfactory and this may be why their suggested framework for unifying penal purposes has been widely rejected.[10]

I argue that punishment is a response to crime. We should understand one in relation to the other. The two are linked and "there can be no just punishment for an unjust crime … Penal justice is linked with just criminalization within a just legal system."[11] Laws are necessary to manage the inevitable conflicts between community members over time. These procedures form a legal system. The criminal law aims at the protection and maintenance of individual legal rights, understood as substantial freedoms worthy of protection for each member based on a political conception of justice.[12] This perspective does not endorse any particular view of justice or freedom, but claims to be consistent with most leading views. The idea is that these individual rights have importance and the criminal law gives effect to this by criminalising theft to honour property rights and so on. Some rights are more central than others. The right to life

Brooks, 'On F. H. Bradley's "Some Remarks on Punishment"', *Ethics* 125 (2014): 223—225.

[10] See Brooks, *Punishment*, chapter 7 and Thom Brooks, *Hegel's Political Philosophy: A Systematic Reading of the Philosophy of Right*, 2d (Edinburgh: Edinburgh University Press, 2013).

[11] Brooks, *Punishment*, 127.

[12] See Brooks, *Punishment*, 127.

has greater importance than the right to property insofar as the former makes possible the latter.

My unified theory of punishment is built on the idea that crimes are rights violations that threaten the substantial freedoms protected by law. Punishment is a response to crime and it aims at the protection of individual legal rights threatened by crime: 'Punishment is about the protection of rights.'[13] The unified theory of punishment is 'unified' because it provides a new framework from which to weigh how different penal purposes can be applied coherently. We consider how these purposes might best contribute to the protection and maintenance of rights. Desert will be crucial, but so will factors like crime reduction and rehabilitation. We must balance them together.

Some rights are more central than others and, likewise, their corresponding punishments will differ, too. The more important the right or need to protect it, the more substantive the necessary response. Crimes like murder should be punished more than theft or larceny because murder is a violation of a more central right. The relative importance of a right will depend on individual circumstances and a changing background context.

One example I give of the unified theory in practice is as *punitive restoration*.[14] This is a reformulated idea of restorative justice considered earlier. Restorative justice proponents are divided (and unconvincing) about what exactly is 'restored'. The unified theory claims we restore rights through their protection and maintenance. Restorative justice claims that possible outcomes should exclude hard treatment. However, if some

[13] Brooks, *Punishment*, 128.

[14] See Brooks, *Punishment*, 132, 136, 142-43 and Thom Brooks, 'Stakeholder Sentencing' in Julian Roberts and Jesper Ryberg (eds), *Popular Punishment: On the Normative Significance of Public Opinion for Penal Theory* (Oxford: Oxford University Press, 2014): 183—203.

forms of intensive hard treatment can be better than other alternatives at helping offenders overcome drug and alcohol dependency or other problems, then these more punitive options might better contribute to their restoration. I discuss several studies that provide evidence to support this view. The claim is not that we should always use punitive options or that greater punitive sentences are desirable, but rather that a restorative justice aiming to restore rights should have such options at its disposable for relevant cases. This can be a way of better embedding restorative justice into the criminal justice system as well: by expanding its options, we might expand its use. This reformulated view of restorative justice I call punitive restoration to draw attention to its being open to more punitive options. More is said about this example of the unified theory of punishment in the rest of the book as I defend the unified theory against its opponents.

Part 3 – Case studies

The last part of the book considers several case studies: capital punishment, juvenile offending, domestic violence and sex crimes. My purpose is to show how different theories of punishment relate to practices and the problems this can lead to.

One example is capital punishment. I argue restorative justice proponents reject hard treatment and so we might suppose reject capital punishment, too. However, standard theories of restorative justice is inapplicable to cases of serious violent crime. So restorative justice might be opposed to the death penalty, but we might accept that it might be justified in some cases because restorative justice is inapplicable. Or we might argue that rehabilitative theories are necessarily opposed to the death

penalty: if someone is executed, then they cannot be rehabilitated. But this is untrue. If someone were incapable of being rehabilitated, then this worry would no longer be relevant and capital punishment might become justified.

A more interesting case is retribution. It is widely thought that retribution is at least always open to justifying capital punishment. While we may have different ideas about what might be deserved, some might claim death is deserved and this raises questions about whether retributivists can oppose the death penalty. I argue they can because they take desert so seriously. If we are unable to say with certainty that someone has desert for a capital offence, then this is a retributivist reason to oppose the death penalty. This is relevant because there are several cases of people convicted and sentenced to death despite having a fair trial and where appeals were exhausted only to have their sentences quashed because of new DNA evidence exonerating them. This shows our judgement can be wrong despite our best efforts. This does not mean retributivists should oppose punishment in any case because of the possibility of making a mistake because most can be remedied, but an execution cannot.[15] The point of my chapter on capital punishment is to make clear that most theories of punishment lack any clear answer for one side or the other. This discussion of retribution is a good example why this is the case.

I also consider juvenile offenders. The punishment of non-adults separately from adults is relatively recent going back to the Illinois Juvenile Court Act of 1899. Youth justice raises interesting questions. Should non-adults be punished any

[15] See also Thom Brooks, 'Retributivist Arguments against Capital Punishment', *Journal of Social Philosophy* 35 (2004): 188—197 and Thom Brooks, 'Retribution and Capital Punishment' in Mark D. White (ed.), *Retributivism: Essays on Theory and Policy* (Oxford: Oxford University Press, 2011): 232—245.

differently from adults? If so, why? And what consequences might there be for the criminal justice more widely? I argue that different theories offer competing reasons for this distinction and which are rarely explored. A retributivist might argue that non-adults are incapable of possessing full moral responsibility for their actions and so would have less desert than if these actions were performed by an adult. But a deterrent proponent might send out different warnings for youths than adults if more effective at generating greater deterrence.

A key issue is age and its relevance. There is normally a distinction of the child where a person is never held responsible for a crime, juvenile offenders who are older than children and below 18 years old, and adults who are 18 or older. I argue that we should instead consider targeting separately the age ranges of 15 to 17, 18 to 24 and leave full adulthood status for 25 years old or more.[16] The reason is that mid-teens represent different types of criminal offenders from 18 to 24 year olds. More effective targeting of specific needs could possibly yield less offending—and less serious offending—at 18 and above.

I discuss the importance of *restorative recognition* through the idea of stakeholding.[17] The many risk factors associated with juvenile offending includes troubled home life, drug and alcohol abuse, peer group pressure and negative support networks. Of course, these factors do not determine offending: many people may be at risk, but nonetheless avoid crime. So what about having these risk factors can make some youths more likely to engage in criminal behaviour?

I argue that stakeholding is key. This about viewing oneself as having a stake in society. Someone who fails to see themselves as

[16] See Brooks, *Punishment*, 182.
[17] See Brooks, *Punishment*, 184.

having a stake is at greater risk. We should focus on reducing incidences of risk factors, but not for their own sake and instead with a view to promoting stakeholding. As I say in the book: 'There is a central need to promote stakeholding and assist young adults in taking greater control over their lives by helping them to see themselves as having a future stake in society'.[18] Reducing risk factors might not be enough. We should aim to reduce risks while promoting a conviction that a young offender does have a stake in the society. This kind of recognition may be difficult, but it is key.

Domestic abuse receives relatively little attention from most theories (and theorists) of punishment. One issue is a question about what kind of crime it is. Several U.S. states have laws criminalising domestic violence, but there can be important differences between how the crime is defined. In contrast, domestic violence has been prosecuted in England and Wales as one or more crimes: 'domestic violence' is not one offence, but a combination of offences. This feature renders it one of the most violent crimes or set of crimes: it may not only consist of sexual and physical abuse, but much more with repeated occurrences.

Again, different theories of punishment can move in different directions. So we might argue a retributivist would argue for a more severe punishment in proportion to the greater violent harms associated with domestic violence. But there are also arguments in favour of informal procedures like restorative justice.[19] This is because some victims do not want their partners imprisoned, but instead want the behaviour to stop. In the book, I favour a middle path and argue for punitive restoration as an option:

[18] Brooks, *Punishment*, 187.

[19] See Linda G. Mills, *Insult to Injury* (Princeton: Princeton University Press, 1999).

> Our choice need not be between prison or non-prison, but perhaps some combination. For example, a brief custodial sentence may serve as a beneficial 'cooling off' period for abusers where they immediately receive some of the therapeutic assistance they require to end their abusive behaviour […] Intensive sentencing is an option that may help offenders most when they are most in need.[20]

Punitive restoration is an option that not all victims will want to explore. But it is an option that has support from some victims and punitive restoration is one way of showing how a unified theory of punishment is possible and preferable bringing together considerations of desert, deterrence, rehabilitation and more with the overall goal of protecting and maintaining rights.

Finally, I consider the case of sexual crimes and, specifically, rape and child sex abuse. *Punishment* explores the different arguments available for their criminalisation and punishment. One issue that arises is a problem for deterrence. Conviction rates are relatively poor and this renders inconclusive what data we have on reoffending. Child sex offenders tend to receive relatively few reconvictions, but they also tend to be found guilty of more prolific crimes when convicted.[21] So if we wanted to punish in order to deter, the reconviction rate is fairly low and might suggest that a more punitive sentence unwarranted although this would receive little public support. My discussion considers a variety of issues concerning how punishment relates to this topic and makes a case for punitive restoration, inspired by work conducted by my Durham University colleague Clare McGlynn.[22]

The book concludes by highlighting the importance of linking the justification of punishment with the justification of its

[20] Brooks, *Punishment*, 197.
[21] See Brooks, *Punsihment*, 205.
[22] See Clare McGlynn, 'Feminism, Rape and the Search for Justice', *Oxford Journal of Legal Studies* 31 (2011): 825—842.

corresponding crime and implications for our viewing 'responsibility as accountability'.[23] Of course, all this rests on a wider view of justice and this is where the book ends. While I do not come out in favour of any particular theory of justice, I claim that whatever it is should be consistent with the idea of a stakeholder society where we each have a stake in our community's life.

Conclusion

Punishment ends:

> Punishment is a topic that never lacks debate. Nothing seems more fitting given the importance of the issues at stake. If you care about justice, then you should care about punishment. This book is an attempt to explain why.[24]

I leave it to readers to judge for themselves how successful I am at achieving this goal.[25]

Durham University

[23] See Brooks, *Punishment*, 215.

[24] Brooks, *Punishment*, 216.

[25] I am very grateful to Gianfranco Pellegrino, Michele Bocchiola, Vittorio Bufacchi, Michele Mangini and Mario Ricciardi for comments and discussion on *Punishment* during my visit to LUISS earlier this year.

If you need to cite this article, please use the following format:

Brooks, Thom, "A précis of *Punishment*," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), 3-23, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

THE PHILOSOPHY OF PUNISHMENT

HOW NOT TO DEFINE PUNISHMENT

BY

ANTONY DUFF

© 2015 – *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015): 25-41
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

How not to Define Punishment

Antony Duff

Brooks offers a critical survey of different normative theories of punishment, finding serious problems with them all, and argues that we should adopt 'the unified theory of punishment' that he draws from Hegel and the English Idealists.[1] I had intended to focus this paper on 'the unified theory', to ask whether it is indeed both genuinely unified and plausible; but I was so taken aback by what Brooks says about the definition of punishment in the early pages of the Introduction that I have focused instead on that. It might seem misguided to devote so much attention to these first few pages: but if one is going to engage in definitional discussion, it is important to get it right.

Much ink, at least some of it wasted, has been spilled on the definition of punishment.[2] Brooks offers this definition [pp. 1-2]:

> (1) Punishment must be for breaking the law.
>
> (2) Punishment must be of a person for breaking the law.
>
> (3) Punishment must be administered and imposed intentionally by an authority with a legal system.
>
> (4) Punishment must involve a loss.

[1] Thom Brooks, *Punishment* (London: Routledge, 2012); all bare page references in the following text are to this book.

[2] For what is still a useful discussion, see D E Scheid, 'Note on Defining 'Punishment'' (1980) 10 *Canadian Journal of Philosophy* 453.

 Philosophy and Public Issues (New Series), Vol. 5, No. 1 (2015): 25-41
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

This definition diverges in some ways from familiar definitions, such as Hart's;[3] but Brooks fails to show that the divergences constitute advantages.

I

Punishment within and outside the Criminal Law

The most striking divergence from other definitions is that Brooks reserves 'punishment' for *criminal* punishment—punishments imposed by a legal authority for the commission of what the law defines as a crime.[4] Other kinds of imposition that we might call 'punishment', 'in our casual everyday talk', should not properly be so called; the reason for this, it seems, is that 'they involve arbitrary executive decisions made by private individuals outside of a legal system'. By contrast, when someone is subjected to criminal punishment she is 'not punished simply because someone else disagreed with her', but 'because of a particular act that she performed' [p. 2]. Now one could indeed argue that, given criminal punishment's distinctive features (the harshness of the sanctions it can involve, its relation to the state and the state's coercive power), a justificatory theory of criminal punishment will need to be different from whatever justificatory theories we might offer of other kinds of punishment—though it may be argued in response that we can find useful connections of meaning between these various practices of punishment; but to

[3] As Brooks notes; see H L A Hart, *Punishment and Responsibility* (2nd ed; Oxford: Oxford University Press, 2008), 4-5.

[4] I leave aside here the question of whether Brooks would reserve 'punishment' for criminal punishment, as imposed by a criminal court for the commission of a criminal offence; or would also allow us to count e.g. 'punitive damages' awarded by a civil court as punishments.

argue that we should not, when we are being 'precise' [p. 2], count anything other than criminal punishment as punishment is a more radical claim, which Brooks fails to justify, since the contrast he draws between criminal punishment and extra-legal 'punishments' is spurious. He is in good company in focusing on criminal punishment, and might also appeal to Hart's limitation of 'the standard or central case of punishment' to punishment imposed by legal authority for offences 'against legal rules';[5] but not even Hart's company can render his arguments persuasive.

First, we might agree that what is imposed arbitrarily, for no good or relevant reason, is not punishment: the punisher must at least claim that there is good reason for the imposition, and that reason must involve the punishee's (alleged) commission of a punishable wrong (I comment later on whether such claims and allegations must be true). However, just the same is true of punishment imposed outside the law. It is true even of the example on which Brooks focuses, that of a parent punishing a child: if what I do to my child is to count as punishment, I must claim that the imposition is justified as a response to some wrong of which the child is guilty—i.e. that there is that good and relevant reason for what I do. It is more obviously true of other kinds of punishment to which Brooks pays less attention. A range of institutions—including schools, universities, religious organisations, many kinds of business, professional associations—operate with codes of ethics or discipline, and with officers or committees who are authorised to impose punishments on those who violate them: what is imposed can count as a punishment only if it is purportedly imposed for the commission of a specified offence, and is imposed by someone with the authority to do so. Parents and disciplinary committees

[5] Hart, n. 3 above, 5; punishments imposed outside the law are 'relegate[d] to the position of sub-standard or secondary cases'.

can of course punish arbitrarily: they can define the norms arbitrarily or retrospectively; they can reach decisions about guilt on inadequate or irrelevant grounds; they can impose punishments whose character and severity are arbitrary. But just the same is true of criminal courts, and of the legislatures that make the laws which the courts apply; such arbitrariness is objectionable, for the same kind of reason, in each case.

Second, some punishments outside the criminal law are imposed by 'private individuals', as in the case of parents, though even there it matters that the parent can claim the authority (legal and moral) to punish. Others, however, are imposed by the authorised officials of the institution whose code is being applied: when a teacher, or a university discipline committee, punishes a student for some misconduct, they are not acting as 'private individuals'. We can agree that punishment involves (a claim to) authority—and that authority will often be either defined or at least constrained by the law; but the authority need not be that of a court of law.

Third, criminal punishment is often, perhaps typically, imposed 'because of a particular act that [the punishee] performed', and not 'simply because someone else disagreed with her' (nor even just because the court disagreed with her, unless the disagreement was about, for instance, whether she had a legally cognizable justification for her admitted commission of an offence). However, first, it is not a definitional feature even of criminal punishment that it must be for an act (even if we take 'act' to include 'omission'): the law can define thought crimes, or crimes of status or condition, for which people can be punished; and whilst those who argue that criminal liability, and thus punishment, *should* be imposed only for or on the basis of an act will object to such laws, their objection is not and could not plausibly be that they are incoherent in authorizing 'punishment'

for something other than an act.[6] Second, in many other punitive contexts, punishments are typically imposed for particular acts specified in the relevant disciplinary code, and are not imposed 'simply because someone … disagreed with' the person being punished: if I am to portray what I impose on you as a punishment, I must claim that it is imposed for a breach of some norm that I have the authority to enforce. Anything that is to count as punishment must be purportedly imposed for an 'offence': but that is not to say either that the offence must be one defined as criminal by the law, or that it must consist in or involve 'a particular act'.

So far, then, we have been given no good reason to reserve 'punishment' for the criminal punishments imposed by legal authorities acting under the aegis of the criminal law. Outside the criminal law, and outside the law, individuals or bodies can claim the authority to impose what they call punishments on those who have broken a relevant code or norm; we have been given no reason to dismiss such calling as merely 'casual everyday talk'.

II

Punishment as Necessarily of Offenders?

Punishment, as careful definers often put it, must be of an alleged offender for an alleged offence.[7] Brooks allows no such qualification in his definition: punishment is 'of a person for

[6] See D N Husak, 'Does Criminal Liability Require an Act?', in Husak, *The Philosophy of Criminal Law* (Oxford: Oxford University Press, 2010), 17; R A Duff, *Answering for Crime: Responsibility and Liability in the Criminal Law* (Oxford: Hart, 2007), ch. 5.

[7] Compare, among others, Hart (n. 3 above), 5.

breaking the law' [p. 1], and—as if to avoid any doubt—'[w]hen we speak of someone being *punished* in this book, we refer to someone who has committed a crime' [p. 2]. On the face of it this seems an odd restriction, since it forbids us to object that punishment is unjust when it is imposed on an innocent person; such impositions, on the Brooks definition, do not count as punishments, and thus cannot be condemned as unjust punishments. Brooks himself seems to ignore this point, when he writes of '[t]he objection … that it is always unjustified to punish those who have not broken the law':

> [w]hen a person is innocent, this person has not acted in such a way that would warrant punishment and, thus, he should be unpunished. [p. 4]

But if punishment is by definition of someone who has committed a crime, the punishment of an innocent is not *unjustified*, or something we *should* not do; it is impossible.[8]

There is a close conceptual connection between punishment and guilt, a connection that reflects a deep normative connection between justified punishment and guilt, and it might be tempting to emphasise the normative connection by presenting it as if it were conceptual:

> Even if the world gathered all its strength, there is one thing it is not able to do, it can no more punish an innocent one than it can put a dead person to death.[9]

[8] See too n. 11 [p. 217]: Brooks tells us that he 'will speak interchangeably of punishment's "definition" and "justification" … because punishment is unjustified where the definitional parts are not fully present'. But if a definitional element is missing, there is then no punishment that could be either justified or unjustified, *as* punishment. There might be (depending on which elements are present) an imposition of some kind; but the mere fact that an imposition does not count as punishment cannot render the imposition unjustified.

It is a mistake, however, simply to conflate the conceptual and the normative connections. The conceptual connection concerns what must be claimed or alleged if an imposition is to count as a punishment: if what I do to *V* is to count as punishing her, I must *claim* that she is guilty of an offence to which this imposition is a response I am authorised to make. The truth, or warrantability, of that claim bears not (directly) on the definitional question of whether what I do is punish *V*, but on the normative question of whether or how what I do is justified.

Consider the two kinds of case in which we might talk (in 'our casual everyday talk') of punishing an innocent person. In one, *V* is deliberately framed by the police or prosecutor, or is convicted by a judge or jury who believe her to be innocent: those who procure this result, being aware of the person's innocence, are deliberately punishing an innocent—although if the imposition is to count as a punishment at all, they must of course claim that she is guilty. In the other kind of case, there is no deliberate miscarriage of justice: *V* is convicted because the lawfully obtained evidence of her guilt left, in the court's honest opinion, no room for any 'reasonable doubt' of her guilt; but, tragically, she is in fact innocent—as might become clear when new evidence later emerges.[10] Now in the first kind of case a convicted innocent might indeed protest that she is not being *punished*—she is being scapegoated, or persecuted. If the claims made about her guilt are obviously spurious, if the scapegoating is manifest, we

[9] S Kierkegaard, *Purity of Heart is to Will One Thing* (trans. D Steere; London: Fontana, 1961), 85.

[10] These cases mark, of course, the two ends of a spectrum; between them fall a range of cases in which there is some more or less serious defect in the way in which an innocent person comes to be convicted (corrupt or careless investigations, evidence that is not properly examined, the drawing of hasty conclusions …), but no deliberate attempt to procure the conviction of a known innocent.

might agree with her. This is not, however, simply because the claim that she is guilty is false: it is because that claim is so manifestly fraudulent that what is being done is obviously the mere pretence of punishment. In the second kind of case, by contrast, we (and she) would be more likely to say that she was indeed being punished—but punished mistakenly and thus unjustly: for in that case the claim of guilt is made reasonably and in good faith.

This reflects a more general point about a range of concepts, of which punishment is one. Sometimes a concept that picks out a particular kind of activity or enterprise includes as part of its meaning the immediate normative criteria for the success or legitimacy of that activity or enterprise: this is true, for instance, of such concepts as education, medicine and (as some would argue) law itself.[11] If a doctor administers a drug to a patient as part of his treatment, and the drug harms the patient rather than curing him, its administration counts as a failure, as bad medicine, by the normative criteria internal to the very idea of medicine: it still counts as medical treatment (at least if it is intended to heal rather than harm); but it must be judged as defective qua medical treatment. If that failure was radical enough, if it would have been obvious to any competent doctor that this drug was not suitable, we might indeed say (by way of rhetorical emphasis) that he was being poisoned rather than treated, or that the doctor was not a real doctor but a quack: if something fails radically enough to satisfy the core normative criteria for being a good *X*, we might say that it is not ('really') an *X* at all. If the doctor is not even aiming to heal rather than harm (if, for instance, she is taking the opportunity to harm the patient under the guise of treating him), we might say that she is not acting qua doctor, is not engaged

[11] See further R A Duff, *Trials and Punishments* (Cambridge: Cambridge University Press, 1986), ch. 3.

(even badly) in medicine—that she is merely pretending to do so: which implies that, as a conceptual matter, medical treatment must be aimed at healing or benefitting the patient. We *might* say the same about punishment (though I don't think that the conceptual issue is so clear here): that if an imposition is to count as punishment at all, not only must it be claimed that it is being imposed on a guilty person, it must be aimed at the guilty; that if it is inflicted on someone who is known or believed to be innocent, it should no longer count as punishment, but only as a pretence of punishment. We might be tempted to say that, but need not decide here whether we should say that, or say rather that it would be flagrantly unjust as punishment.[12] All we need note here is that even if punishment must definitionally be aimed at the guilty, and must be of the actually guilty if it is to be justified as punishment, it counts as punishment (albeit necessarily as punishment that fails as punishment) even if it misses that aim—even if it is mistakenly imposed on an innocent.

It might seem that I have laboured this point unnecessarily: for I have agreed with Brooks that there is a conceptual or definitional connection between punishment and guilt, and that if the person on whom some hardship is inflicted is actually innocent, then what is inflicted on her cannot be justified as punishment; so why should it matter whether we express the point by saying (as I would) that it is punishment, but unjustified as such; or by saying (as Brooks would) that it is not punishment (properly speaking)? It matters partly because we should be clear about the difference, and the connections, between definitional

[12] The institutional character of criminal punishment might be relevant here: if the system as a whole is aimed at punishing only the guilty, its occasional abuse by individuals to procure the 'punishment' of victims they know to be innocent might still count, in virtue of its institutional context and character, as punishment.

and justificatory claims, and about the different ways in which normative criteria might be involved in the criteria for the correct application of a concept. But it matters too because if we are concerned, as Brooks is, with the justification of criminal punishment, we are concerned with a practice that is, like any human practice, unavoidably fallible: it will inevitably sometimes convict and punish an innocent person, however earnestly those working within it try to avoid such errors; if we are going to punish anyone, we will sometimes punish an innocent. A justification of a system of criminal punishment must thus be a justification not of a system that only punishes the guilty, but of a system that punishes only those who are found guilty through a fair criminal process; of a system, that is, that sometimes punishes the innocent.

III

Punishment and Loss

Brooks rightly resists [p. 5] the suggestion that punishment must include pain; instead, he suggests, we should define it in terms of 'loss'.[13] But what is not clear from his definition is whether that loss must be intended as a loss. Punishment must be 'imposed intentionally', and 'must involve a loss' [p. 2]; but that leaves open the possibility that the loss could be a foreseen, but not intended, aspect of the imposition. When a tort defendant loses his case and is ordered to pay damages to the plaintiff, that

[13] I'd rather say that punishment must be burdensome, to emphasise that punishment can be undertaken willingly by a repentant offender: whilst I can willingly undertake or embrace a burden, as a burden, I cannot do more than accept a loss as a necessary cost of something else—I cannot embrace it as a *loss*.

payment involves a loss (though the loss might be relatively painless, if the amount is small relative to his means); but ordinary damages are not understood as punishment, even if they are awarded by a legal authority against a person who has broken the law in the sense that he has failed to take the care that, according to the law, he ought to take.[14]

It might seem that, on a plausible reading of Brooks' definition, the loss must be intended as a loss: for if punishment 'must involve a loss', as part of its meaning, to intend to impose punishment must be to intend to impose loss; if I intend what I do to count as punishment, I must intend it to involve a loss. A later comment, however, suggests otherwise.

> Suppose there is a violent psychopath. He is genuinely suffering from psychopathic delusions that compel him to attempt killing innocent persons without provocation. He lacks culpability for his actions, but these actions present a clear danger to the public. The unified theory of punishment might argue that the violent psychopath should be incapacitated regardless of culpability. [pp. 140-1]

Now if this person's incapacitation is to be justified by a theory of *punishment*, it presumably must be justified precisely as a punishment; but is that how we should understand this kind of incapacitation? One puzzle is that it is not clear what sort of case Brooks has in mind, since 'psychopathy' is not normally understood as involving 'delusions' which might 'compel' the person to violence; that is one reason why there is continuing controversy about the criminal responsibility, and culpability, of

[14] That is of course why 'punitive damages' are distinguished from ordinary damages: they are intended not, like ordinary damages, to provide compensation for the harm suffered by the plaintiff, but rather to burden the defendant.

psychopaths.[15] So a theory of punishment might well justify the punishment of a violent psychopath, on the simple grounds that he is culpably guilty. But if we instead think of someone who lacks culpability, because his actions were 'compel[ed]' by his delusions (someone who would count as psychotic rather than as psychopathic), then it is not clear why we should count his detention as a punishment. He would, if he came to trial at all, be a strong candidate for the insanity defence—rightly so, if he lacks culpability. The court would no doubt order his detention in a psychiatric institution, to protect both him and others; but it would not do so as a criminal punishment for his crime—at least as we normally understand the idea of punishment. The key point here seems to be that if our aim is simply to incapacitate someone who is radically disordered and non-culpable, in order to protect others, it surely should not be our *intention* to inflict any loss on him. The only sufficiently secure method of incapacitation might in fact involve a loss; but if we could sufficiently incapacitate without inflicting loss, we should do so; and we should try to minimise whatever loss cannot be avoided. This is, of course, a standard way of distinguishing the compulsory detention of the mentally disordered from the punishment of culpably responsible offenders: the latter, but not the former, must be intended to be burdensome (or to cause a loss, in Brooks' terms). But it seems that Brooks would count all such detentions as punishments, if the detained person has committed a crime, since all involve a loss (of freedom).[16]

[15] For some useful recent readings, see L Malatesti and J McMillan (eds.), *Responsibility and Psychopathy* (Oxford: Oxford University Press, 2010).

[16] There might be room for argument about whether the incapacitative detention of a non-culpable, disordered offender is imposed *for* his crime: but that is not something that seems open to Brooks, given what he says about the deluded, non-culpable 'violent psychopath'.

There are two ways in which Brooks could deal with this issue. One—the more orthodox way—would be to distinguish punishment from other modes of crime-related state coercion by holding that punishment must be intended to cause loss (or to be burdensome); in which case a theory of punishment, unified or not, will have nothing to say about the incapacitative detention of the non-culpably dangerous (except that it is not punishment). The other would be to expand the definition of punishment to cover such coercive practices as the detention of the non-culpably dangerous; but then we would need to know how far that expansion should go. In particular, if the focus is on preventing the dangerous from harming others, why should the commission of a crime be a condition for detention: it might be evidentially significant, but if we could be confident that they are dangerous on the basis of other evidence, why wait until they kill? That would not count as 'punishment' in 'our casual everyday talk'; but if what we justifiably do to the 'violent psychopath' to prevent his further crimes is to count as justified punishment, it is not clear why punishment in its now expanded sense should require an (alleged) offender.

My own view is that we should, for the sake of both analytic and normative clarity, stick with the narrower definition of criminal punishment, as an intentionally burdensome response to one who has been convicted of committing a crime: that is a distinctive practice, different in its normative character from other coercive practices such as the detention of those judged to be in some way dangerous; if it is to be justified at all, it requires a distinctive normative rationale. I suspect that this is also Brooks' view: but then he needs, first, to make clear that punishment must be intended to involve a loss; and second, to rethink his comments about the 'violent psychopath'.

IV

Punishment and Expression

Another way in which some theorists would distinguish punishment from other kinds of coercive imposition is by arguing that punishment involves, definitionally, the expression or communication of censure or condemnation.[17] When we punish an offender, the burden we impose is not *merely* a burden; it carries, and is intended to convey, a message about what he did. (Such a definitional claim is not yet a normative claim about the proper justifying aims of punishment: it is, rather, a claim about just what it is that needs to be justified if we are to justify a system of punishment.)

Brooks notes, and gives short shrift to, the definitional claim [pp. 3-4].[18] Now the claim is certainly arguable—but his rejection of it is radically under-argued. He finds in Feinberg's article an identification of 'punishment' with imprisonment, other species of non-custodial sanction being classed as 'penalties': but, as he notes, most convicted offenders receive non-custodial sentences; a theory of punishment that counted only imprisonment as punishment would not be a theory of punishment as we practise it. Feinberg does sometimes appear to identify punishment (as expressive), or the 'hard treatment' that constitutes punishment, with imprisonment, though elsewhere he is more careful to recognise that imprisonment is just one form of punishment.

[17] See, for different versions of this idea, J Feinberg, 'The Expressive Function of Punishment', in Feinberg, *Doing and Deserving* (Princeton NJ: Princeton University Press, 1970) 95; I Primoratz, 'Punishment as Language' (1989) 64 *Philosophy* 187; A von Hirsch, *Censure and Sanctions* (Oxford: Oxford University Press, 1993); R A Duff, *Punishment, Communication and Community* (New York: Oxford University Press, 2001).

[18] He pays more attention later (ch. 6) to normative theories of punishment as expression or communication.

Penal theorists also sometimes talk as if criminal punishment consists in and only in imprisonment (or capital punishment). But, first, other theorists, including some who take censure-communication to be a defining feature of punishment, make it clear that criminal punishment need not be custodial, and argue that we should make less use of prison than we do.[19] Second, the fact that both criminal punishment, and what Feinberg and others count as 'penalties' rather than 'punishments', can consist in a fine does not by itself go any way towards showing that we cannot usefully distinguish punishments, as communicating censure, from penalties as mere sanctions. Brooks argues that

> [t]he view that penalties and 'punishments' (understood as imprisonment) are different in character is … a distinction drawn too sharply that we should reject. [p. 3]

However, if punishment is 'understood as imprisonment', it is clearly different in character from non-custodial penalties: the distinction between custodial and non-custodial sanctions can be drawn quite sharply and clearly.[20] More importantly, if we understand punishment to encompass a range of non-custodial sanctions, we can still draw a clear distinction between censure-communicating punishment and penalties that lack such a communicative dimension.

Monetary sanctions provide the simplest example here. An official requirement to pay a specified sum of money could constitute any of a variety of kinds of imposition: it could be a tax demand; an award of damages, or an order to pay a sum owed, arising from a civil case; an administrative penalty for breach of a

[19] See e.g. von Hirsch, n. 17 above; Duff, n. 17 above.

[20] Though some glossing would be needed to deal with such phenomena as the suspended prison sentence, or with cases in which breaching the requirements of a non-custodial sanction can attract a prison sentence.

non-criminal regulation; or a criminal punishment imposed following conviction. What the requirement amounts to, what it means, depends on the grounds on which it is made; on the institutional context in which it is made; and on the terms or tones (themselves determined partly by the institutional context) in which it is made. On the definitional suggestion under discussion here, it should count as a punishment only if it is intended to convey a formal censure of the conduct because of which the requirement is made.

Such a distinction is formally drawn, for instance, in German law, which distinguishes crimes (*Straftaten*) from regulatory infractions (*Ordnungswidrigkeiten*), the latter being dealt with under a regulatory code that is separate from the criminal code. Monetary sanctions are available under both codes—as criminal fines (*Geldstrafen*) for *Straftaten*, as administrative penalties (*Geldbussen*) for *Ordnungswidrigkeiten*. The meaning of the monetary sanction is, however, different in the two cases: for crimes attract a formal condemnation expressed in the sanction—the fine is imposed for conduct that is reproachable (*vorwerfbar*) and blameworthy (*schuldig*), whereas administrative penalties lack such a censorial meaning.[21] I do not suggest that the distinction is unproblematic:[22] but it is not a manifestly untenable distinction; and one could not accuse German law of confusing punishment with imprisonment.

There is room for argument about whether it is useful to include censure-communication in our definition of (criminal)

[21] See *Gesetz über Ordnungswidrigkeiten* (1968; consolidated in 1975); for a useful (and critical) introduction see T Weigend, 'The Legal and Practical Problems Posed by the Difference between Criminal Law and Administrative Penal Law' (1988) 59 *Revue Internationale de Droit Pénal* 67.

[22] See the doubts raised by the European Court of Human Rights in e.g. *Öztürk v Germany* (1984) 6 EHRR 409; also Weigend, n. 21 above.

punishment, to mark out the particular kind of practice that is to be theorised; but Brooks' selective critique of Feinberg does not give us reason to doubt the possibility, or utility, of distinguishing punishments from penalties.

V

Finally

It is not clear how much time it is useful to spend on the definition of punishment (or of criminal punishment). We do need to mark out the particular practice, or range of practices, that we aim to subject to normative theorising; but that is not to say that we need to offer a 'definition' of punishment. However, if we are going to offer a definition, as Brooks does at the start of his book, we need to do so with care and attention to detail—something that I fear Brooks has failed to do.

University of Minnesota &

University of Stirling

If you need to cite this article, please use the following format:

Duff, Antony, "How not to Define Punishment," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), 25-41, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

THE PHILOSOPHY OF PUNISHMENT

PUNISHMENT AND COHERENCE

BY

MICHELLE MADDEN DEMPSEY

 Philosophy and Public Issues (New Series), Vol. 5, No. 1 (2015): 43-56
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Punishment Theory and Coherence

Michelle Madden Dempsey

Thom Brooks is to be commended for having taken on an almost impossible task in writing his impressive new book, *Punishment*.[1] His stated goals are ambitious: "to present a critical guide to the latest research on the leading theories of punishment and the most important alternative approaches…to consider their application in particular contexts, such as the use of capital punishment, juvenile offending, and the punishment of domestic violence, rape, and child sex offences … and to present the most thorough explanation and defence … to date" of his own contribution to the punishment literature, which he coins the "unified theory of punishment" (ix-x).

What is all the more impressive is that Brooks aims to accomplish these goals whilst writing in a style that is accessible to a general, non-specialist audience, thus avoiding what Americans would characterize as an "inside-baseball" approach to punishment theory. While Brooks' efforts are mostly successful, the book does at times suffer from a lack of clarity and thoroughness.

In what follows, I will press on areas where Brooks' *Punishment* might have benefited from further argument. This discussion is offered not so much as a critique of the book Brooks has written,

[1] Thom Brooks, *Punishment* (London: Routledge, 2012). Page numbers in the text are to this book.

 Philosophy and Public Issues (New Series), Vol. 5, No. 1 (2015): 43-56
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

but as an invitation to further address these underdeveloped areas in future work.

I

What Counts as a Successful Theory of Punishment?

In order to judge the success of Brooks' unified theory of punishment, or indeed any theory of punishment, we should begin with an account of what a theory of punishment is and what criteria are appropriate to evaluating its success. Brooks is more or less clear about what a theory of punishment is. Following H.L.A. Hart, he argues that a theory of punishment consists of three parts: a definition of punishment, an identification of the "general justifying aim" of punishment (or, in the case of hybrid theories, the "general justifying aims" of punishment), and an account of how punishment should be distributed (6).

With respect to his definition of punishment, Brooks again follows Hart in stipulating a definition that limits his inquiry to legally imposed punishment. The only explanation offered for this limitation is the counterintuitive view that "[i]t would be unacceptable for any individual to act in a private capacity in carrying out punishments" (5). Yet, there are many instances of private punishment that are entirely justified, such as a parent giving a "time-out" to a child who hits her younger sibling. We should not assume away the existence and justification of non-

legal punishments – nor should we presuppose that legal punishment presents the central case of punishment.[2]

With respect to the "general justifying aims" of punishment, Brooks is clear that his unified theory embraces a plurality of penal goals. "Punishment need not be either retributivist, deterrent, or rehabilitative, but all at once" (211). While I am sympathetic to Brooks' pluralist approach to justifying punishment, it seems wrong to frame the issue in terms of "aims" or "goals" of punishment. Rather, we should formulate the point in terms of what punishment actually *does*, not merely what it *aims* to do. A system of punishment that aims to deter but never actually manages to deter is unjustified from the perspective of deterrence theory. A system of punishment that aims to impose deserved punishment on the guilty but only ever manages to punish the innocent is unjustified from the perspective of retributive theory. A system of punishment that aims to express condemnation and/or shape social norms but conducts its activities in secret is unjustified from the perspective of an expressive theory. And so on. I do not mean to detract from the justificatory work that can be done by having valuable aims.[3] My point here is simply that what punishment actually does matters as well, and that framing the issue in terms of "general justifying aims" obscures this point. We should, perhaps, instead frame the inquiry in terms of "reasons for punishment."

[2] See, John Gardner, *Introduction*, HLA Hart, *Punishment and Responsibility* (OUP 2nd ed 2007), p. liii. Although, as discussed below, that legal punishment is inflicted by the state, on behalf of the community – rather than on behalf of the victim – is an important aspect of its displacement value. See n 14.

[3] As I've explained in regard to the justification of prosecutorial action, trying can have value (which I referred to as "telic value") even when the chances of success are remote. Michelle Madden Dempsey, *Prosecuting Domestic Violence: A Philosophical Analysis* (Oxford: Oxford University Press 2009) p. 65.

Brooks has comparatively little to say on the issue of distribution as a theoretical matter, but his practical applications of the unified theory in the later third of the book offers insights to his views regarding the distribution of punishment. To his credit, Brooks resists drawing too sharp a distinction between concerns of general justification and distribution, recognizing instead that "[p]erhaps the consequences should matter" even to the question of distribution (97-98).

What remains somewhat unclear throughout the book, however, are the criteria we should use to evaluate whether any given theory of punishment is successful. We might suppose that the criteria for a successful theory of punishment simply tracks the elements of what a theory of punishment is. If so, then a successful theory of punishment will:

(1) identify salient features of punishment (thus providing a successful definition of punishment);

(2) illuminate considerations relevant to the justification of punishment as a general practice (thus successfully identifying its "general justifying reasons"); and

(3) specify the conditions under which punishment may be justified in a particular instance (thereby providing a successful account of how punishment should be distributed).

Are there any further criteria in determining what constitutes a successful theory of punishment? For Brooks, the answer seems to be a resounding yes: a successful theory of punishment must be *coherent*. Indeed Brooks is concerned throughout to emphasize the coherence of his unified theory of punishment: emphasizing repeatedly that "[t]he unified theory of punishment is a unique attempt to bring together several different principles of punishment within a single and coherent approach" (123).

Unfortunately, however, it is not clear what coherence means in this context, or why it should be regarded as a necessary feature of a successful theory of punishment. If a successful theory of punishment is meant to provide an account of whether, how and why punishment is justified, then it seems sufficient to point out any and all salient features that count in favor of punishing, either as a general practice or in a particular case. We should expect these features to vary from society to society, from crime to crime, and from case to case.[4] In some instances, the justifications may resemble one another. In other instances, the justifications may bear little resemblance. If this is so, perhaps we should agree with John Gardner:

> [Criminal punishment is] such an extraordinary abomination, that it patently needs all the justificatory help it can get. If we believe it should remain a fixture in our legal and political system, we cannot afford to dispense with or disdain any of the various things, however modest and localized, which can be said in its favour.[5]

Let us refer to theories of punishment that ascribe to this view as "Pick-a-Mix" theories of punishment, and distinguish them from what we might call "Coherence" theories. Pick-a-Mix theories stake no claim regarding whether any institution or particular instance of punishment is justified; they simply observe that if we wish to keep punishing, then our justification for so doing should be based on any and every consideration that

[4] Indeed, Brooks seems to acknowledge as much when he observes that "political societies may … punish the same crimes in the same individual circumstances very differently in some part due to possible differences in societal contexts. Crimes and punishments may significantly differ from one political society to the next" (137).

[5] John Gardner, *Offences and Defences* (Oxford: Oxford University Press 2007), p. 214.

weighs in its favor.[6] If something can be said in favor of punishment, then (according to the Pick-a-Mix theories) we should include that consideration in our justification of punishment. This is true even of considerations that weigh only very weakly in favor of punishment, and of considerations that weigh in favor of only some kinds of punishments but not others. It is also true of considerations that do not manifest any degree of theoretical coherence. For the Pick-a-Mix theorist, worrying about the degree of coherence amongst considerations that bear on the justification of punishment is simply a waste of intellectual energy. We should instead dedicate ourselves either to identifying salient features of punishment that genuinely weigh in its favor, or set ourselves the task of abolishing punishment for lack of justification.

Brooks rejects this approach to punishment theory, illustrating his disdain through a detailed criticism on the Model Penal Code sentencing principles, which simply list multiple goals of punishment without attending to concerns of how these various considerations cohere.[7] Brooks' complaint against Pick-a-Mix theories of punishment (as illustrated in the Model Penal Code) is not that it draws upon multiple penal goals. Indeed, his unified

[6] Gardner, n 5, p. 214.

[7] The Model Penal Code approach is embodied in §1.02:

(2) The general purposes of the provisions governing the sentencing and treatment of offenders are:

a. to prevent the commission of offences;

b. to promote the correction and rehabilitation of offenders;

c. to safeguard offenders against excessive, disproportionate or arbitrary punishment;

d. to give fair warning of the nature of the sentences that may be imposed on convictions of an offence;

e. to differentiate offenders with a view to a just individualization in their treatment.

theory of punishment is similarly pluralistic. The problem, Brooks argues, is that simply placing "multiple penal goals [in] a list lack[s] a suitably robust framework that offers a sufficiently clear steer on how these goals relate to one another within the framework" (132). He explains:

> Why should any of these goals be included? The answer seems to be that each is intuitively attractive on its individual merits. But this fails to address specifically how each might relate. Imagine making a cake combining only those ingredients that you enjoy individually. Following this procedure may not guarantee that all the necessary ingredients for making a cake are included. Nor is there any guarantee that the cake will be edible. Now imagine starting a company by inviting only those persons that you enjoy working with individually. This procedure may not guarantee that all the necessary tasks will be covered. Nor is there any guarantee that the company's members will work together suitably effectively. These examples centre on the problem of justifying a legal practice without sufficient consideration of how the individual parts coherently work together in support of the practice aims (132-133).

But what is the special ingredient in Brooks' theory that makes the plurality of penal goals he endorses cohere any better than the laundry list of reasons on offer in a Pick-a-Mix theory? What is it that makes the unified theory unified? Brooks' explanation is opaque. He claims that "[t]he unified theory of punishment overcomes this problem" of incoherence because "[i]t addresses desert, proportionality, and other penal goals [as] they come together within a larger unified framework" (133). To this point in his explanation, we must take it on trust. The unified theory is unified because Brooks keeps telling us it is.

Yet, how does this unity manifest itself? How does the coherence of the unified theory inform the way we think about punishment, so that our thinking is different, better than it would be under a Pick-a-Mix theory? Brooks offers the following response: "[Under a unified theory approach] we don't weigh up

possible sentences in light of general deterrence versus desert and other penal considerations because we find them intuitively attractive individually" (133). I confess to not understanding what Brooks means at this point. Does he mean that we don't weigh up possible sentences in light of general deterrence versus desert and other penal considerations at all – or merely that we don't do so because we find them intuitively attractive individually?

The remainder of Brooks' explanation does little to clarify the point: "Punishment does not bring together multiple penal goals because it can, but because it should. Punishment is a response to crime that aims at the restoration of rights. Punishment addresses multiple penal goals in serving its aims" (133). While I have no disagreement with any of these claims, it remains unclear how they explain what it means to say that the unified theory is coherent in a way that makes its penal pluralism more attractive than other hybrid theories of punishment.

Any plausible explanation would have to point to one or more penal goals (reasons for punishment) that play a cohering role in the theory's account of how and when punishment is justified, and explain how each goal hangs together in a coherent whole. Brooks offers no account of how this cohering relation between multiple penal goals is achieved under the unified theory. Yet, there is such an explanation available to another hybrid theory that Brooks rejects: expressivism.

II

The Coherence of Expressivist Theories of Punishment

An expressivist theory of punishment can provide the coherent penal pluralism Brooks prizes in the following way.

First, the theory will identify multiple operative reasons for punishment similar to the list that Brooks endorses: retribution, deterrence, etc.[8] All are operative reasons for punishment insofar as they are values that can be realized through punishment - each operative reason is capable of doing some normative work in justifying punishment. Second, the expressive theory will identify how the expressive function of punishment provides auxiliary reasons for punishment that relate to each of operative reasons.[9] The expressive function of punishment provides auxiliary reasons because it identifies punishment as an act which there is reason to perform under the circumstances. Which is to say, even if the expressive function of punishment is neither a complete reason, nor even an operative reason for punishment, the expressive function of punishment nonetheless helps to identify punishment as a justified response to the defendant's crime for reasons of deterrence, retribution, etc.—and as such the expressive function of punishment serves as an auxiliary reason for punishment.[10]

[8] On operative reasons, see Joseph Raz, *Practical Reason and Norms* (Oxford: Oxford University Press 1990), pp. 33-34.

[9] On auxiliary reasons, see Raz, *Practical Reason and Norms*, pp. 34-35. Raz describes two roles auxiliary reasons play in practical reasoning: identifying and strengthening. Identifying auxiliary reasons "help identify the act which there is reason to perform."

[10] I take no view here as to whether the expressive function of punishment is also an operative reason – that is, whether there is a value in expressing whatever it is that punishment expresses. In previous work, I have argued that such value exists and grounds operative reasons for prosecutors to pursue (or not pursue) certain kinds of prosecutions. My point here is that even if the expressive function of punishment does not have value, it does help identify punishment as the means for realizing other values such as deterrence, retribution, etc. – that is, it functions as an auxiliary reason of the identifying type. Raz, *Practical Reason and Norms*, pp. 34-35. The expressive function of punishment may also serve as an auxiliary reason of the strength-affecting type. I assume as much in the discussion in the main text.

As the expressivist theory we are examining is pluralistic, it can admit of multiple operative reasons for punishment – deterrence-based operative reasons, retributive-based operative reasons, displacement-based operative reasons, etc.[11] The relationship between these various operative reasons can be illuminated by explaining how each relates to the expressive function of punishment: specifically, the fact that punishment is expressed/communicated as the intentional infliction of a loss *for breaking the law* (that is, its expressive function) is an auxiliary reason that picks out punishment as a particularly effective way to realize deterrent, retributive, and displacement value. I will consider each relationship in turn below.

The expressive function of punishment serves as an auxiliary reason relating to deterrence-based reasons in favor of punishment, because it helps to identify punishment as a particularly effective way to realize the value of deterrence. As many have argued and as Brooks agrees, deterrence is indeed an important value that can be realized by punishment – which is to say, deterrence is an operative reason in favor of punishment. Yet, the deterrent effect of punishment depends to a significant degree on the fact that the punishment is *expressed* to the defendant and potential future defendants as the intentional infliction of a loss for breaking the law. Imagine a punishment that is not communicated to the defendant. The state imposes a fine in response to the defendant's crime, but it does so not by public declaration of the punishment, but instead by simply withdrawing the funds from the defendant's bank account or sneaking into his home to obtain the cash. Absent the public expression of the punishment, the defendant is likely to be confused, not deterred. Similarly, such a punishment would be

[11] On displacement, see Gardner, *Offences and Defences*, pp. 213-216, and the text accompanying nn 13-16 below.

incapable of achieving general deterrence. Unless the punishment is communicated to potential future defendants as the intentional infliction of a loss for breaking the law, they are far less likely to be deterred by punishment. In this way, the expressive function of punishment serves as the handmaid of a deterrence-based justification of punishment.

Now consider retribution-based operative reasons in favor of punishment and how they relate to the expressive function of punishment. Retributive theories are correct in supposing that there is value in a defendant suffering some intentionally inflicted loss for having committed a crime.[12] Which is to say, retributive value is something that can be said in favor of punishment. Yet, if punishment lacks its expressive function, then the defendant will not register the value of suffering *for his crime*. If the state imposes a secret punishment, making the defendant's life go less well in any variety of ways, but never communicates to the defendant that this treatment is being imposed *for his crime*, then the retributive value of the punishment not realized. Just as with deterrence, the expressive function of punishment serves as the handmaid of retributive justifications of punishment.

There is a similar story to be told with respect to displacement-based operative reasons in favor of punishment. As John Gardner puts it:

> That people are inclined to retaliate against those who wrong them, often with good excuse but rarely with adequate justification, creates a rational

[12] I would rather put the point in terms of "for having committed a wrong that also happens to be a crime" to screen out cases in which a defendant is punished for having committed a crime that is not also a wrong. Brooks frames the point in terms of crime, so I will follow suit here.

> pressure for social practices which tend to take the heat out of the situation and remove some of the temptation to retaliate.[13]

This "rational pressure" is what I have referred to as displacement-based operative reasons in favor of punishment. Its displacement value is indeed something that can be said in favor of punishment. Yet, again, the expressive function of punishment is key to realizing punishment's displacement value. Here, we should expand our understanding of the expressive function of punishment beyond the characterization offered above, and layer in the fact that *legal punishment* is imposed by the *state* – acting on behalf of the community – rather than being imposed by victims or on behalf of victims.[14] By imposing legal punishment, the state (community) expresses to potentially vengeful victims something along the lines of, "Just chill … we've got this. We will address this crime adequately – so that you will have no reason to take matters into your own hands."[15] To the extent that the state fails to express that *it alone* is the primary agent in inflicting punishment, the displacement value of punishment is reduced. Moreover, to the extent that the state fails to make good on its promise of an adequate response to crime, it fails to displace reasons victims may have to take matters into their own hands. Unfortunately, the criminal justice system often stumbles on both

[13] Gardner, *Offences and Defences*, p. 214.

[14] Above (n 2), I noted that Brooks was wrong to suppose that legal punishment was the central case of punishment. Still, if a key pillar of the justification of legal punishment lies in its displacement value, then the fact that the state imposes punishment is an important feature of the justification of *legal* punishment.

[15] Some criminal law abolitionists have argued that the state stepping in to "steal" victim's conflicts in this way counts against legal punishment. Nils Christie, *Conflicts as Property*, 17 *British Journal Of Criminology* 1-15 (1977). Yet, to the extent that the expressive function of punishment helps to realize a displacement value of punishment, the abolitionist argument is weakened.

fronts: it too often turns its discretionary authority over to victims and fails to provide an adequate response to crime.[16] Still, if a system of punishment is functioning properly, the expressive function of punishment serves as the handmaid of displacement justifications of punishment.

A similar explanation may be offered with respect to other values realized by punishment, but I hope to have done enough to motivate the possibility that an expressivist theory of punishment is capable of providing the coherent penal pluralism Brooks prizes. In sum, such a theory not only explains how punishment can be justified in terms of a plurality of operative reasons in favor of punishment, it illustrates how the expressive function of punishment serves an auxiliary reason that identifies punishment as a particularly effective means of realizing the values that ground those operative reasons.

Moreover, an expressivist theory is attractive insofar as it informs our practices of punishment by highlighting the importance of transparency and publicity in our punitive practices. If we sacrifice transparency and publicity (that is, if we compromise the expressive function of punishment), it becomes all the more difficult to justify our punitive practices. Interestingly, this point holds true even when our punitive practices are beyond reproach – even when we are only punishing the deserving, and only for serious wrongs that cause substantial harm. Often we think of transparency and publicity as important only for uncovering official corruption or discovering and checking the misguided exercise of official discretion. Yet, as the expressivist theory I've outlined above demonstrates,

[16] These failures are most starkly illustrated in domestic violence prosecutions, where victims' stated wishes are often treated as authoritative, while the state nonetheless fails to provide an adequate response to these crimes. See, Dempsey, *Prosecuting Domestic Violence*, ch. 9.

transparency and publicity are central to the justification of punishment even for an otherwise perfectly well-functioning system.

III

Conclusion

In conclusion, Thom Brooks' *Punishment* provides an intriguing and insightful account of punishment and its justification. I find his theory of punishment particularly appealing in virtue of its embrace of penal pluralism. My reflections have focused upon Brooks' search for coherence amongst the plurality of reasons that may weigh in favor of punishment. If the concerns I've raised are correct, this search for coherence may be unnecessary (that is, perhaps the Pick-a-Mix theorists are correct to think that searching for such coherenceis a waste of rational energies). If, however, coherence *is* an important aspect of a pluralistic justification of punishment, then Brooks may do well to reconsider whether the expressive function of punishment can provide that coherence, by unifying otherwise disparate, pluralistic reasons for punishment through its role as an auxiliary reason.

Villanova University

If you need to cite this article, please use the following format:

Dempsey, Michelle Madden, "Punishment and Coherence," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), 43-56, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

THE PHILOSOPHY OF PUNISHMENT

ELABORATING NEGATIVE RETRIBUTIVISM

BY

RICHARD L. LIPPKE

© 2015 – *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015): 57-71
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Elaborating Negative Retributivism

Richard L. Lippke

Thom Brooks' smartly-argued overview of punishment theory and some of its practical implications provides an excellent opportunity to think further about some issues with which I have wrestled previously. In particular, I have come to believe that some version of negative retributivism might be the most defensible approach to legal punishment's justification. Negative retributivism is an intuitively plausible theory because it incorporates both crime reduction and retributive elements into a unified approach. One question, as Brooks notes, is whether this unity is more apparent than real.[1] Can it really be possible to bring together into a coherent theory the "forward-looking" elements of a crime reduction approach with the "backward-looking" elements of a retributive one? A second question, as Brooks also notes, concerns how to understand the retributive constraints on our efforts to reduce future offending that the theory incorporates. How do these constraints amount to anything more than arbitrary stipulations, taken on by crime reductionists in order to craft a theory that is more palatable?[2] Brooks is skeptical about negative retributivism, preferring instead a "unified theory" that conceives of legal punishment's task as the "protection and restoration" of rights.[3] My aim in the discussion that follows is

1 Thom Brooks, *Punishment* (London: Routledge, 2012), p. 98.
2 Ibid., p. 99.
3 Ibid., p. 131.

 Philosophy and Public Issues (New Series), Vol. 5, No. 1 (2015): 57-71
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

both to elaborate negative retributivism and suggest that it offers a clearer and more compelling approach to the justification of legal punishment than Brooks' unified theory.

The discussion is divided into three sections. In the first, I clarify what negative retributivism is, arguing that Brooks has to some extent mischaracterized it. In the second section, I address the two challenges to the theory that Brooks poses: In effect, these come to the question whether it is a coherent theory or a patched together set of more or less arbitrary stipulations about legal punishment. In the third section, I briefly consider Brooks' unified theory, contending both that Brooks' account is underdeveloped and that negative retributivism can be usefully conceived as instrumental in securing a system of legal rights.

I

What Negative Retributivism Is

As it is standardly formulated, negative retributivism posits that the general justifying aim of legal punishment is the reduction of crime, whether such reduction is effected by deterrence, incapacitation, or rehabilitation. However, negative retributivism insists that the pursuit of that aim is to be limited by two retributive constraints, one forbidding the intentional or knowing punishment of the innocent, the other forbidding disproportionate punishment of the guilty. According to negative retributivism, as contrasted with positive retributivism, we need not punish the guilty (if doing so would not, for some reason, reduce crime) or proportionately punish the guilty (again, if doing so would not, for some reason, reduce crime). It is undeserved punishment that negative retributivism rules out; deserved or

(fully) proportional punishment is not required by it. Brooks' characterization of negative retributivism includes the first constraint, but misses the second, as he claims that the theory permits disproportionate punishment of the guilty if crime reduction considerations so dictate.[4] This is mistaken, or so I contend. Negative retributivism is premised on an awareness of the pitfalls of a simple crime reduction approach to legal punishment's justification. It has been argued that such an approach permits both punishment of the innocent and over-punishment of the guilty. The two retributive constraints have been thought necessary to address these defects. Recently, I argued for a third constraint, one forbidding degrading punishment of the guilty.[5] I come back to this in the next section.

Brooks also suggests a rule-utilitarian grounding of the retributive constraints.[6] This, in turn, leads him to question whether they should be seen as operating in cases in which adhering to them would not produce the best overall crime reduction consequences. Why not punish the innocent, or over-punish the guilty, if doing so would, in a given set of circumstances, reduce future offending more than other measures that we might take? However, the retributive limits on legal punishment that negative retributivism incorporates are more commonly conceived as deontological side-constraints. Thus, they rather more firmly forbid punishment of the innocent or excessive punishment of the guilty, even if doing so would reduce crime. Yet as Brooks might point out, this raises the question of what sort of grounding, if any, these constraints have. We might also wonder how "firm" they should be conceived to be.

4 Ibid., p. 33.

5 Richard L. Lippke, "Some Surprising Implications of Negative Retributivism," *Journal of Applied Philosophy* 31 (2014): pp. 49-62, at 55.

6 Brooks *Punishment*, p. 98.

II

Addressing Brooks' Challenges

Is there a plausible account of the retributive constraints, or are they little more than stipulations taken on by crime reductionists desperate to save their theory? I believe that there is such an account, though I also believe that its elaboration might produce a more retributively-flavored theory of legal punishment than negative retributivism is often conceived to be.

To begin with, we should not intentionally or knowingly punish the innocent because they do not deserve legal punishment's characteristic censure and hard treatment, neither of which make sense unless imposed on individuals who are capable of morally responsible action and have engaged in seriously wrongful and therefore proscribed conduct.[7] The innocent have not violated the criminal law, or if they have done so, they lack the kind of culpability that makes their conduct worthy of legal punishment's condemnation.[8] Also, legal punishment of those who have violated the criminal law ought to reflect the gravity of their wrongs.[9] If it does not, then it is not responsive to them as individuals, censuring their conduct in ways that are appropriate given the nature and extent of their wrongs.

Anticipating Brooks own unified theory of punishment, with its grounding in a scheme of justified legal rights, we might say that the innocent have acted in ways consonant with their legal rights and have thereby respected, or at least deferred to, the

[7] For the notion that legal punishment involves censure and hard treatment, see Andrew von Hirsch, *Censure and Sanctions* (Oxford: Clarendon Press, 1993).
[8] Retributivists standardly recognize a variety of exemptions, excuses, and justifications which shield from penal sanctions those who have technically violated the criminal law.
[9] See von Hirsch, *Censure and Sanctions*, p. 15.

equal rights of others. Or, if the innocent have infringed upon others' rights, they have done so accidentally, justifiably, or excusably and so do not merit punishment. It would seem deeply incoherent to set up a system of legal rights, with the aim of seeing to it that all citizens can fully and equally enjoy them, and then to devise institutions of legal punishment that are indifferent to, or worse, scornful of, whether individuals have remained within the bounds of their rights and so behaved responsibly. Of course, it might be wondered why we should not, on occasion, infringe the rights of the innocent, by legal punishment, in order to deter future right violations of more numerous or grave kinds.[10] But negative retributivism should be conceived as forbidding trade-offs among citizens' rights of this kind, because they would allocate censure and hard treatment in ways that is unresponsive to the conduct of individuals. For similar reasons, negative retributivism forbids excessive punishment of the guilty, even if it would better deter future right violations. Excessive censure and hard treatment are also unresponsive to the character of the misconduct of offenders, not taking proper account the ways and extent to which they infringed others' rights and allocating blame accordingly.

However, there are hard cases for negative retributivists and these suggest that its constraints need not be conceived as absolute. For instance, Antony Duff and Stephen Morse have both argued, though in different ways, that preventive detention of "dangerous" individuals past the point of their deserved punishment for previous offenses might be warranted in some

[10] Such an approach would be one which Robert Nozick once referred to as involving "utilitarianism of rights." See his *Anarchy, State, and Utopia* (New York: Basic Books, 1974), p. 28.

cases.[11] Such detention can be conceived as deliberate punishment of the "innocent," in the sense of confining individuals when they do not deserve it. Or it can be conceived as disproportionate punishment of the guilty—above the level at which they deserve it for their past crimes. In either case, if the arguments of Duff or Morse are convincing, they suggest that the retributive constraints might be construed as presumptive only, albeit strongly so. Importantly, the trade-off of rights contemplated in such cases is responsive to the conduct of the individuals to be detained in ways that do not run afoul of the retributive insistence that punishment be imposed only on the blameworthy. This is because there might be individuals whose past actions strongly suggest that they have enduring dispositions towards violence, such that we might believe that we have little choice but to abridge their rights in order to protect the rights of others. This, in turn, suggests that the retributive constraints, though not entirely inflexible, can be rebutted only by certain kinds of considerations. Moreover, it could be argued that the "dangerous" individuals whose rights are thereby abridged ought to be compensated for their losses or at least that the conditions under which they are detained ought to be made as non-punitive as possible.[12]

Beyond the two standard retributive constraints, a third one seems needed. Given the ways in which the legal punishment of individuals makes sense only against a conceptual and normative

[11] R. A. Duff, "Dangerousness and Citizenship," in A. Ashworth and M. Wasik (eds.), *Fundamentals of Sentencing Theory: Essays in Honor of Andrew von Hirsch* (Oxford: Clarendon Press, 1998), 141-63, and Stephen J. Morse, "Blame and Danger: An Essay on Preventive Detention," *Boston University Law Review* 76 (1996), pp. 112-55.

[12] For doubts about the feasibility of non-punitive detention, see my "No Easy Way Out: Dangerous Offenders and Preventive Detention," *Law & Philosophy* 27 (2008), pp. 383-414.

backdrop according to which they are conceived as capable of morally responsible action, legal punishment should not be allowed to turn them into beings of another kind, thereby degrading them in the process. In Jeffrie Murphy's memorable words, we should not turn those punished into "terrified, defecating, urinating, screaming animals."[13] Punishment that does so cannot be rationally engaged with by its unfortunate recipients as expressing appropriate public opprobrium of their conduct; neither can it be taken to heart by them and thereby used as an impetus to moral self-improvement. Beyond the obvious cases of torture or mutilation, sanctions such as prolonged solitary confinement or imprisonment under radically corrupt or insecure conditions, can make it nearly impossible for offenders to retain and exercise their moral capacities. Even more ordinary imprisonment of the wrong kinds can degrade, especially if prolonged.[14] Like the more familiar retributive constraints, the non-degradation constraint can be cast negatively; it need not entail devising legal sanctions designed to positively promote the moral capacities of offenders. Nonetheless, a well-informed and thoughtful approach to crime reduction must acknowledge both the cost and limited efficacy of legal punishment, and thus recognize the advantages of convincing and enabling offenders to refrain from further misconduct all on their own. This points us in the direction, I believe, of raising the profile of rehabilitation within the crime reduction element of negative retributivism. Put simply, we will see less crime in the future if offenders are not

[13] Jeffrie G. Murphy, "Cruel and Unusual Punishments," in M. A. Stewart (ed.), *Law, Morality, and Rights* (Dordrecht: D. Reidel, 1979), pp. 373-404, at 387.

[14] See, for instance, Craig Haney's disturbing account of the effects of supermax confinement in his "Mental Health Issues in Long-Term and 'Supermax' Confinement," *Crime & Delinquency* 49 (2003), 1pp. 24-56.

degraded (as the retributive constraint enjoins) but also prodded and helped to be more morally responsible.[15]

Assuming that the preceding provides an intelligible backstory for the retributive constraints, how should we conceive their normative character within a theory of legal punishment? As noted earlier, Brooks posits a rule-utilitarian grounding for them, which he sees as fragile. I have suggested that they are better construed as deontological side-constraints on legal punishment, though perhaps not absolute ones in all cases. But are they ever absolute? And if they are presumptive only, then what kinds of considerations will rebut the presumptions and how strong must those considerations have to be? I can do little more here than adumbrate answers to these important questions.

Even if punishment's censuring and hard treatment components must be keyed to the nature and severity of the legal wrongs individuals have committed, it is apparent that the concept of proportionality in punishment is a somewhat loose constraint on the sanction scale, unless sanctions become either cardinally or ordinally disproportionate in the extreme. [16] Very harsh sanctions for minor offenses might be absolutely ruled out by desert considerations. Shoplifters, for instance, should not be punished with life sentences, no matter if doing so would optimally reduce crime. Similarly, though somewhat more controversially, very mild sanctions for serious offenses seem to sleight the victims of crimes and thus fail, absolutely as it were, to censure offenders proportionally.[17] Murderers should not be

[15] See Lippke, "Some Surprising Implications of Negative Retributivism," p. 57.

[16] For the distinction between cardinal and ordinal proportionality, see von Hirsch, *Censure and Sanctions*, p. 18.

[17] Some theorists broadly sympathetic to negative retributivism argue that it should be conceived as placing only upper limits on sanctions, not as requiring

punished with modest fines, or the rights of their victims will be diminished to the point of insignificance.

When it comes to ordinal proportionality, again, the constraint seems close to absolute when the comparative injustices are extreme. Murderers should be punished more harshly than shoplifters come what may, or else punishment will utterly fail to be apportioned to blame. Yet as we move away from clear instances of either cardinal or ordinal disproportionality, matters become murkier. It is widely agreed that there are no uniquely proportional sentences for the various crime types.[18] At most, considerations of cardinal and ordinal proportionality might yield sentence ranges for each of the various different crime types, within a sentencing scheme that ranks crime types from less to more serious. Such a scheme could permit some overlap of the sentence ranges for higher and lower-ranked crimes types, in recognition of the possibility that some crime tokens will be unusual in the harms they produce or the culpability with which agents acted.[19]

Offense tokens should be assigned sentences within the relevant ranges. So long as the ranges are not permitted to become too large, like offenses will be sentenced roughly alike, in accordance with a plausible ordinal proportionality constraint. In determining sentences for offense tokens, we could conceive of desert considerations as being weakly presumptive. This would mean that like offenders ought to receive like sentences, unless

some minimal level of punishment for crimes. On this, see Richard S. Frase, *Just Sentencing: Principles and Procedures for a Workable System* (New York: Oxford University Press, 2013), pp. 25-31.

[18] See R. A. Duff, *Punishment, Communication, and Community* (Oxford: Oxford University Press, 2001), pp. 133-34, and Jesper Ryberg, *The Ethics of Proportionate Punishment: A Critical Investigation* (Dordrecht: Kluwer Academic, 2004), p. 131.

[19] Frase, *Just Sentencing*, pp. 48-49.

clear and empirically validated crime reduction considerations support moving them upward or downward within the relevant sentencing range.[20] To move the sentence for a crime token out of the relevant range, we might insist that only very powerful considerations would suffice, such as those cited previously in cases of offenders whose persistent violent dispositions have been amply demonstrated.

The non-degradation constraint seems near absolute in strength when it comes to the more extreme ways in which punishment practices can destroy or defeat the moral capacities of offenders. Proscriptions against torture or mutilation ought to be stringent; the same seems true of bans on extended solitary confinement, slow starvation or exposure to constant physical brutality or insecurity, or the failure to treat prisoners' medical conditions. Again, if legal punishment is to communicate censure to offenders, which is to be understood and taken to heart by them, then sanctions which disable moral understanding or render it moot must be avoided, more or less at all costs. There are other forms of hard treatment that will reduce crime without resorting to such sanctions. Still, legal sanctions can be somewhat at odds with preserving the moral capacities of offenders without being starkly so, or might be at odds with doing so only if they persist over the fairly long term. And there might be reasons to employ sanctions of these kinds that are sufficiently compelling in some cases. For instance, there might be circumstances in which it makes sense to isolate offenders from the rest of the prison population for some period of time—to discipline them or foil

[20] Frase argues that recidivist premiums—that is longer sentences for offenders with previous criminal histories—might be justified in some cases because empirically validated. However, such premiums should not be typically permitted to move an offender's sentence for her most recent offense out of the relevant sentencing range. See Frase, *Just Sentencing*, Chapter 4.

repeated escape attempts by them—even if doing so is in tension with the non-degradation constraint. This, in turn, suggests that the constraint is presumptive only when the losses or deprivations imposed by legal punishment are less blatantly degrading. Still, the authorities should have to prove that the conduct of the individual in question is the basis for the further infringement of his rights when it risks degradation. Also, the constraint might be held to leave a trace, requiring the authorities to undertake positive measures to ameliorate the impact of whatever losses or deprivations are needed to protect the rights of others.

III

Brooks' Unified Theory

I believe that Brooks is right to seek a theory that both (a) incorporates as many of the standard justifying aims of legal punishment as can be coherently accommodated, and (b) does so in a principled—as opposed to patched-together—fashion. I say this as someone who has over the years defended positive retributivism but who has gradually come to the conclusion that crime reduction plays an ineliminable role in legal punishment's justification. It is not the censuring or communicative component of legal punishment that gives me pause, but the hard treatment component. As von Hirsch wisely pointed out some years ago, it seems difficult to account for that component without conceding that its role is partly to discourage future offending or render individuals less capable of it for some period of time.[21]

[21] von Hirsch *Censure and Sanctions*, p. 12. Of course, there have been attempts to explain the hard treatment element of legal punishment according to

Importantly, conceding a crime reduction element to legal punishment need not render us vulnerable to Hegel's charge that a deterrent element to legal punishment means we are simply threatening people, as a man "lifts his stick to a dog."[22] As von Hirsch noted, the censuring aspect of legal punishment can be conceived as communicating with us as beings capable of understanding the justified character of legal constraints on our conduct and the importance of observing them. At the same time, we can acknowledge the powerful internal and external forces that incline us towards wrongdoing, and thus the need to provide ourselves with a (not disproportionate) prudential incentive back-up to shore up our resolve to behave responsibly, even if this might mean that some of us will abide by the law for purely prudential reasons.[23]

I also believe that Brooks is on the right track in pointing to a theory of human rights and the protection of such rights within a legal scheme as providing some of the conceptual and normative backdrop for a theory of legal punishment. However, I am less convinced than he appears to be that an account of human rights can be quickly or easily mined for a complete theory of legal punishment. Numerous approaches to legal punishment seem consistent with its serving a scheme of legally defined and protected human rights.

Brooks' unified theory is premised on the notion that legal punishment "protects and restores" the legal rights violated or infringed by criminal offending. Yet Brooks does not spend

retributive logic. See, for instance, John Kleinig, "Punishment and Moral Seriousness," *Israel Law Review* 25 (1992), pp. 401-21, and R. A. Duff, *Trials and Punishments* (Cambridge: Cambridge University Press, 1986), pp. 240-54.

[22] G. W. F. Hegel, *The Philosophy of Right*, T. M. Knox (trans.), (Oxford: Oxford University Press, 1965), p. 246.

[23] von Hirsch, *Censure and Sanctions*, p. 13.

enough time unpacking this crucial notion. There are at least three different ways in which legal punishment might be held to "protect and restore" rights: First, it could require those who have violated others' rights to provide them with restitution, in an attempt to make them whole again. Such an account faces various difficulties, among them the fact that many crimes violate or infringe no one's rights directly, yet seemingly must still be punished. Also, it is civil law, more than criminal law, which seems concerned with making the victims of irresponsible conduct whole again. Indeed, some rights violations seem appropriate dealt with entirely by the civil law (e.g., violations of contracts). Further, there is the risk that well-off offenders will commit crimes and pay restitution with relative ease. This threatens to blur the important distinction between crimes, which involve wrongful conduct, and torts, which involve "priced" but not outright prohibited conduct.[24]

Second, legal punishment could "protect and restore" rights through its censuring aspect, condemning violators and, more positively, communicating emphatically the importance of rights to both offenders and other citizens. As previously noted, whether such an account can explain the characteristic hard treatment aspect of legal punishment is less clear. It is also unclear whether, without hard treatment, such an account would provide much crime reduction, unless the idea is that by communicating the importance of rights, we shore up the collective resolve to abide by them.

Third, legal punishment could "protect and restore" rights by assuring members of the law-abiding public that those who refuse to abide by reasonable restrictions on their conduct will not get

[24] John C. Coffee, Jr., "Paradigms Lost: The Blurring of the Criminal and Civil Law Models—and What Can Be Done About It," *Yale Law Journal* 101 (1992): pp. 1875-93.

away with it, but will instead suffer censure and hard treatment.[25] No society can tolerate for long law-breaking in its more serious forms, ones that violate fundamental rights, without becoming de-stabilized. A scheme of rights will be imperiled if fearful citizens engage in anticipatory violence or predation, or if angry or vengeful ones believe that they have to take matters into their own hands in dealing with lawbreakers. When the criminal law is reliably enforced, incentives to abide by it are in place and the law-abiding can see what happens to those who are indifferent to or defy the law's strictures.

In his chapter on unified theory, Brooks appears to embrace all of these accounts of how legal punishment can "protect and restore" rights. Yet I think it is fair to say that he does not spell out carefully how these accounts are to be integrated into a coherent whole. Simply repeating that legal punishment aims at the "protection and restoration" of rights, as he too often does, leaves in place all of the problems that would have to be addressed in articulating a unified theory. Also, it is worth noting that legal punishment, as Brooks admits, curtails or infringes the rights of offenders, and imperils the exercise and enjoyment of them by ex-offenders through its intensely stigmatizing and disabling qualities. Hence, to say that legal punishment's role is to "protect and restore" rights requires more by way of a complex and nuanced explanation than Brooks provides.

I believe that a suitably developed form of negative retributivism might be of service to Brooks in his endeavor to articulate a unified theory. We should seek to secure human rights not only by communicating their importance, but also by threatening and enacting punishment of individuals who violate laws designed to

[25] Consider in this regard the notion of "dominion" as it is developed and defended by John Braithwaite and Philip Pettit in *Not Just Deserts: A Republican Theory of Criminal Justice* (Oxford: Clarendon Press, 1990): pp. 64-65.

protect the enjoyment and exercise of rights. But we must take stringent measures to ensure that those we punish have indeed violated the law, can be held legally accountable for doing so (because they were morally responsible beings at the time who behaved unacceptably), and are not punished disproportionately to their wrongs. We must also ensure that they are punished in ways congruent with their coming to understand and taking to heart the censure communicated by legal punishment, so that they can work to bring their future conduct in line with defensible legal restrictions.

Indiana University

If you need to cite this article, please use the following format:

Lippke, Richard L., "Elaborating Negative Retributivism," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), 57-71, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

THE PHILOSOPHY OF PUNISHMENT

DEFENDING *PUNISHMENT*

REPLIES TO CRITICS

BY

THOM BROOKS

© 2015 – *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015): 73-94
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Defending *Punishment*
Replies to Critics

Thom Brooks

I am very grateful to the contributors for this symposium for their essays on my *Punishment* book. Each focuses with different elements of my work. Antony Duff examines the definition of punishment in my first few pages.[1] Michelle Madden Dempsey analyses the importance given to coherence in my account and critique of expressivist theories of punishment.[2] Richard Lippke considers my statements about negative retributivism in an important new defence of that approach.[3] I examine each of these in turn below. While I do not change my position, they draw attention to certain features in my overall argument worth reflecting on at greater length. So I welcome this opportunity to address and clarify these now and grateful for their helping me to rethink my original arguments.

[1] See R. A. Duff, How not to Define Punishment," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), pp. 25-41.
[2] See Michelle Madden Dempsey, "Punishment and Coherence," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), pp. 43-56.
[3] Richard L. Lippke, "Elaborating Negative Retributivism," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), pp. 57-71.

 Philosophy and Public Issues (New Series), Vol. 5, No. 1 (2015): 73-94
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

I

Duff on Definitions

Duff begins the symposium challenging the definition of punishment that starts my book, citing my proposed definition:

> (1) Punishment must be for breaking the law.
>
> (2) Punishment must be of a person for breaking the law.
>
> (3) Punishment must be administered and imposed intentionally by an authority with a legal system.
>
> (4) Punishment must involve a loss.[4]

My purpose is to define and clarify what is meant by the term 'punishment' in my book. This definition should make clear that my use of 'punishment' is restricted to the breaking of law by individuals administered and imposed intentionally by an authority involving a loss within a legal system. So my aim is to consider punishment as *a legal practice* and examine its justification.

This aspect is important. Part of my argument is that too many discussions about punishment fail to connect punishment with crime. It is true we often hear talk about 'punishing' a child for misbehaviour, but I argue this talk is metaphorical and that such a practice is different from our legal practices—and these legal practices are my focus. Either there is nothing distinctive about 'legal punishment' versus talk of punishment in other contexts, or this difference matters and I claim that it does.

Duff first denies that punishment must be for breaking the law. He says:

> A range of institutions—including schools, universities, religious organisations, many kinds of business, professional associations—operate with codes of ethics or discipline, and with officers or committees who are

[4] Thom Brooks, *Punishment* (London: Routledge, 2012), pp. 1-2.

> authorised to impose punishments on those who violate them: what is imposed can count as a punishment only if it is purportedly imposed for the commission of a specified offence, and is imposed by someone with the authority to do so.[5]

At first glance, readers might think Duff and I agree: punishments are only imposed where someone has committed an offence. But notice how Duff makes this point about *punishments* by changing what is meant by *offences*: Duff's reference to 'a specified offence' is to some breach of a code of ethics and not crime. It is hardly surprising that Duff rejects my narrower focus as he counts as an 'offence' more than unlawful conduct and counts as 'punishment' more than actions connected to unlawful conduct. His understanding of possible crimes and punishments is over-inclusive and goes beyond the criminal law and sentencing policy. He refers to 'many other punitive contexts' and their 'disciplinary code' leading him to claim we need not consider as offences conduct that is 'defined as criminal by the law'.[6] Duff's non-legal understanding of offences and their punishment is intended to demonstrate that my narrower focus on criminal law and sentencing is incorrect, but all Duff does here is use one definition to refute another.[7]

[5] Duff, "How not to Define Punishment", p. 27.

[6] Ibid., pp. 28-29.

[7] Dempsey is also critical of this part of my definition stating that 'we should not assume away the existence and justification of *non-legal punishments*—nor should we presuppose that *legal punishment* presents the central case of *punishment*' (emphasis added). This distinguishes between 'punishment' as a category that includes 'legal punishments' and 'non-legal punishments'. In Dempsey's language, my project is concerned entirely with legal punishment (which I refer to as 'punishment'). I don't consider how (legal) punishment might connect with other forms of non-legal sanction: my examination considers the justifications on offer for legal punishment to gain greater clarity within this narrow focus. I do not see how my examination of legal punishment benefits as a project concerned with legal punishment by

Duff next claims that 'careful definers' of punishment note it must be of an alleged offender for alleged offences.[8] He disagrees with my statement that punishment is 'of a person for breaking the law'.[9] Duff claims it is 'an odd restriction' because it demands that punishment be justified and 'it forbids us to object that punishment is unjust when it is posed on an innocent person; such impositions, on the Brooks definition, do not count as punishments' and so cannot be condemned as such.[10] Duff claims we should distinguish between whether what we do to another is punishment and whether it is justified.

But this is an odd criticism. We don't punish people alleged to have committed a crime, but persons convicted for it. Curiously, Duff appears to argue that something counts as punishment if its definition is aimed at the guilty 'and must be of the actual guilty' even where the person punished is innocent, but wrongly sentenced. This is odd because it commits Duff to accepting that (positive) retributivists—that require offenders possess desert in order to justify punishment—would claim that any wrongfully convicted persons are punished despite their innocence. Desert does not only justify the amount of punishment to be distributed, but the distribution itself. Perhaps our disagreement is that Duff calls imprisoning innocent people a form of unjust punishment and I would call it a miscarriage of justice: punishment would be not merely normatively inadequate, but should never have

considering other cases of non-legal sanctions. So I don't doubt that people refer to non-legal practices as punishment (such as punishing a child) and I don't claim they are unimportant or uninteresting, but they are concerns that appear to go beyond the particular phenomena of legal punishment that is my focus. This dispute seems more a quibble over definitions than concerns about substance as far as this specific issue is concerned.

[8] Duff, "How not to Define Punishment", p. 29.

[9] Brooks, *Punishment*, p. 3.

[10] Duff, "How not to Define Punishment", p. 30.

happened. We can agree innocent people endure some form of loss perhaps, but my point remains: punishments are not to be understood or justified isolated from the offences that give rise to them—so this important link between crime and punishment is absent where the innocent are concerned. The criminal justice system does indeed send innocent people to prisons, but they are neither deserved, rehabilitated, etc. because what they endure is not punishment but injustice. And this gives rise to justified rights to make claims for compensation in recognition they did not receive justice.

Duff considers my comments on punishment and loss. He is critical of my brief note that a violent psychopath tempted to kill without provocation might be incapacitated on my unified theory of punishment 'regardless of culpability'.[11] Duff initially states concerns about we should count someone's detention as punishment where they lack culpability. Of course, someone need not be culpable to be convicted of a criminal offence. Examples include possession offences of strict liability.

Duff overlooks a key point. In this part of my book, I was arguing that the unified theory of punishment that I defend takes a distinctive view about the relation between crime and punishment. I argue that the crimes should be understood as violations of rights and punishments is an attempt to restore them. In some cases no such restoration may be necessary and this is one way pardons might be justified on my view. But if punishment is about maintaining a system of rights where crimes are punished in proportion to their centrality within this wider system, then what to make of cases where clear public dangers exist but may lack culpability? My point is that culpability may not

[11] Brooks, *Punishment*, p. 141.

be required to justify the distribution of punishment, including (but not restricted to) cases like this.

Finally, Duff provides a narrow criticism of my fairly extensive rejection of expressivist and communicative theories, including his own theory. Duff focuses on my discussion of Feinberg's distinction between punishment and a penalty where punishment refers to hard treatment such as prison and penalty refers to sanctions. Duff claims this distinction is important and can be made where a sanction 'is intended to convey a formal censure'—and this is true of both hard treatment and 'non-custodial' sanctions.[12]

But this attempted defence concedes my argument. I argue that Feinberg's distinction between punishment as hard treatment and penalties as other forms of sanctions is drawn too sharply because the expression of public censure can be present in sanctions other than imprisonment. I argue this might even be true with verbal warnings. Duff now appears to accept my criticism, but his reason for continuing to see a clear distinction anyway is at best unclear. Moreover, Duff overlooks a key point in my argument that punishments in practice rarely take the form of a prison sentence *or* a monetary fine *or* some other sanction. Instead, two or more might be imposed together as *the punishment* of an offender: so actual court outcomes for an offender can include a combination of a fine, suspended sentence, community order and perhaps others. Our choice is not hard treatment or an alternative, but often which package of penal options are justified for an offender. I argued it was difficult to see how some, but not all, parts of the same punishment could rest on different justificatory bases between expressivist and non-expressivist forms. This line is drawn too sharp because any (justified)

[12] Duff, "How not to Define Punishment", p. 40.

punishment expresses public censure for illegal conduct although each may differ in degree, at least metaphorically and perhaps only metaphorically. But this is a mistake that could have been avoided if legal punishment was more closely tied to the criminal law and sentencing policy.

II

Dempsey on Coherence and Expressivism

Dempsey raises two main concerns with *Punishment*. First, she is critical of the role and importance of coherence in my account of punishment. She rightly notes that I would reject a 'Pick-a-Mix' theory of punishment where we simply select any consideration for justifying punishment that we favour or reject punishment altogether for its lack of justification.[13] Dempsey notes that my criticism of the Model Penal Code is that it is a kind of Pick-a-Mix theory. The Model Penal Code says at §1.02:

> (2) The general purposes of the provisions governing the sentencing and treatment of offenders are:
>
> a. to prevent the commission of offences;
>
> b. to promote the correction and rehabilitation of offenders;
>
> c. to safeguard offenders against excessive, disproportionate or arbitrary punishment;
>
> d. to give fair warning of the nature of the sentences that may be imposed on convictions of an offence;
>
> e. to differentiate offenders with a view to a just individualization in their treatment.

[13] Dempsey, "Punishment Theory and Coherence", p. 47.

The Model Penal Code is a kind of Pick-a-Mix 'theory' of punishment because it offers multiple penal purposes which may clash with one another and without any structure for how any potential clashes can be managed, if not avoided. Moreover, the penal purposes listed in the Model Penal Code may be commendable, but why *these* particular purposes? How should they be considered when applied to particular cases? Missing is a justification of these parts to punishment's justification as a whole.

Dempsey does not disagree with my critique *per se*, but rather my alternative. She says:

> What is it that makes the unified theory unified? Brooks' explanation is opaque. He claims that "[t]he unified theory of punishment overcomes this problem" of incoherence because "[i]t addresses desert, proportionality, and other penal goals [as] they come together within a larger framework." To this point in his explanation, we must take it on trust. The unified theory is unified because Brooks keeps telling us it is.[14]

She concludes: 'Brooks offers no account of how this cohering relation between multiple penal goals is achieved under the unified theory'.[15] For Dempsey, there appears little, if any, substantive difference between Pick-a-Mix theories like the Model Penal Code and my unified theory of punishment.

It is worth reconsidering how the unified theory is unified. Recall the importance of the link between crime and punishment for my account: there is no justified punishment for an unjustified crime. I claim that crimes should be understood as a kind of rights violation. Punishment is justified for the restoration and maintenance of rights. Desert can captured by the importance that someone has violated, for example. Following Alan Brudner, I argue this view of 'legal retributivism' overcomes problems

[14] Dempsey, "Punishment Theory and Coherence", p. 49.
[15] Ibid, p. 50.

found with Legal Moralism's 'moral retributivism'.[16] Penal principles such as crime reduction or rehabilitation can be justified insofar as they can contribute to the restoration and maintenance of rights threatened by crime. Proportionality is determined by considering the centrality of the right affected.[17] Dempsey rightly notes that this view of proportionality concedes that some communities will view the relation between crimes and punishments differently from others. For the unified theory of punishment, this is not problematic *per se* and perhaps inevitable. It may also help us understand how society's set their punishments as an indication for how those who set them view their corresponding crimes with potentially interesting implications over time that I do not consider.

Let me use an example to illustrate, such as theft. This offence is a violation of another's right to possess property. The amount of justified punishment for the thief depends on a consideration of which possible outcomes are most likely to yield best the restoration and maintenance of rights. Outcomes may not be exclusively preventative or rehabilitative: the reformed offender may wish to avoid the threat of the state imposing further rehabilitation costs in addition to his recognising he should avoid such activities anyway. And it is the case that some communities will choose more punitive outcomes than others, but the unified theory attempts an explanation: these differences can be justified because the context matters. A community under threat because

[16] See Alan Brudner, *Punishment and Freedom* (Oxford: Oxford University Press, 2009).

[17] Duff notes my work claims links with Hegel and the English Idealists, but this is somewhat inaccurate because I explicitly connect ideas to the wider British Idealism tradition including Scottish philosophers, such as James Seth and John Stuart Mackenzie. See Brooks, *Punishment*, pp. 127, 129-130, 236-238, 241 and Thom Brooks, "James Seth on Natural Law and Legal Theory", *Collingwood and British Idealism Studies* 12 (2012): pp. 115-132.

of invasion or civil war is likely to become more threatened by criminal acts like theft than other communities enjoying a secure peace. This is not relativism, but *contexualism* (if it should have a name) because context matters. We can avoid a narrow preoccupation with whether one aim *versus* another is satisfied where we can view them more like a toolbox to help us achieve a restoration of rights. This gives theoretical coherence to why *these* aims or purposes should be included (answer: because they can help us achieve our goal of restoring and protecting rights), but unlikely to provide any specific determination of precisely which package of possible outcomes should be decided. But this is no more a problem for the unified theory of punishment than alternatives, where they run into problems of how much might be 'deserved' or what punishment will likely sufficiently deter.

Dempsey's second concern is that expressivist theories of punishment can give me the unified coherence I'm after and a better alternative. Punishment as the expression of public censure 'is an auxiliary reason that picks out punishment as a particularly effective way to realize deterrent, rehabilitative, and displacement value'.[18] Dempsey claims that understanding punishment as expressivist sends a message to offender and, as a message *to* offenders, is thought to communicate some deterrent value. The idea seems to be that if a message is not communicated expressly to a particular individual then it might lack a deterrence effect. I'm unsure about this. Nor do I see that this is how deterrence is more effective, and not what I call macrodeterrence (general deterrence) or microdeterrence (specific deterrence) modes. Dempsey further claims that expressivism captures retributive values in communicating a punishment as *'for his crime'* to offenders.[19]

[18] Dempsey, "Punishment Theory and Coherence", p. 52.
[19] Dempsey, "Punishment Theory and Coherence", p. 53.

I have two concerns with this proposal. The first is whether expressivism is a hybrid theory, in fact. This is considered in chapter 6 of my book and not substantively addressed here (or by Duff who is the principle target of my critique). Expressivism may claim to achieve multiple penal purposes, but they aim to satisfy only one. No expressivist argues that any offender should be punished any more than deserved. It is not implausible to imagine a scenario where an offender who has committed an especially notorious, well publicised crime would receive a lesser sentence if punished for only what is deserved than receive the full brunt of vivid public anger. This causes a particular difficulty for expressivists because they commit themselves to the importance of the public's communication of displeasure while only supporting punishments that meet a different test of retributivist desert. And so I argue in *Punishment* that expressivists—to quote Duff—hold the view that punishment 'must…be understood in retributive terms'.[20]

My second concern is whether expressivist theories of punishment are even theories of punishment. This is because if public condemnation is what matters, then public condemnation might justify any range of outcomes that may have more to do with who people are or represent than what they have done. Again, expressivists seem to fall back on retributivist justifications and it remains unclear what distinctive difference public displeasure brings to our thinking about punishment where it is held that the only permissible penal outcomes must be deserved.

Dempsey claims expressivism can help provide me with the unified theory I am looking for. But there are questions about expressivism's genuine distinctiveness in practice and whether it even is the hybrid theory it presents itself to be. One illustration

[20] See Brooks, *Punishment*, p. 115 and R. A Duff, "Crime, Prohibition and Punishment", *Journal of Applied Philosophy* 19 (2002): pp. 97-108, at 106.

of this is Duff's discussion of punishment as secular penance. What is said to be distinctive about Duff's view is that punishment is not only a matter of we, the public, expressing our condemnation of a criminal act in sentencing an offender, but punishment is *also* a matter of the offender communicating to we, the public, an apology through serving a prison sentence. This second part about communication is what makes the view a communicative theory of punishment and not merely an expressivist theory. But offenders need not do anything at all beyond serve the prison sentence they are compelled to endure by the state. It is bewildering to me how it can be claimed secular penance is happening in communicating some message to the public where the offender is coerced and may not, in fact, communicate or express anything at all.[21] So I am not yet persuaded expressivist theories of punishment are the answer.

III

Lippke on Negative Retributivism

In *Punishment*, I target the idea of *positive* retribution understood as the view that desert is necessary and sufficient for punishment. If an offender can be found to deserve punishment, then this is sufficient to distribute punishment to him. I claim this 'standard view' of retribution is part of 'a rich, venerable tradition' that includes a variety of different ideas about how retribution might be understood.[22]

[21] See Brooks, *Punishment*, pp. 104-105 for this part of my discussion of this view.
[22] See Brooks, *Punishment*, pp. 15, 33.

While positive retribution understands desert as necessary and sufficient for punishment, *negative* retribution sees desert as necessary, but not sufficient: 'the *severity* of punishment may be determined by factors beyond desert, such as favourable consequences'.[23] In my discussion, I note that 'both [positive and negative] retributivisms might endorse similar punishments, but with different justifications'.[24] They each might punish the same offender differently, but I do not say or suggest that either would punish a thief more than a murderer.[25] Lippke claims that negative retributivism has two constraints: the first forbids punishing the innocent and the second forbids 'disproportionate' punishment of the guilty. Lippke says my characterisation captures the first, but not the second although it should also be clear that *nothing* I say about negative retributivism contravenes the second constraint either.[26]

My critique of negative retribution argues that it is a type of rule utilitarianism, 'and perhaps with all the concerns that rule utilitarianism attracts'.[27] The main concern is 'that the justification for the rules that constrain desired consequences may differ from the justification for why we should pursue these consequences'.[28] For example, if desert is so important for selecting who might be punished, why should it not play the most important, if not only, role in determining the punishment's amount? Or if non-desert factors are so important that they should play the most prominent role, then why be constrained by desert if it inhibited pursuit of such non-desert factors? In *Punishment*, I argue that 'perhaps there is good reason to distribute punishment in a

[23] Brooks, *Punishment*, p. 33.
[24] Ibid.
[25] See Brooks, *Punishment*, pp. 33-34.
[26] Lippke, "Elaborating Negative Retributivism", p. 58.
[27] Brooks, *Punishment*, p. 98.
[28] Ibid.

particular way and a different good reason to justify the practice of punishment. What we require is some third reason to justify how these reasons come together, if negative retributivism is to be a theoretically coherent theory of punishment'.[29] My conclusion is that negative retributivist accounts have lacked this theoretical coherence.

Lippke's negative retributivism claims the general justifying aim of legal punishment is crime reduction, but subject to the retributivist constraints concerning we only punish the guilty and not disproportionately so.[30] So how important are non-retributivist factors? We require retributivist desert because it is necessary for justified punishment on this view. But any justified punishment must also be proportionate—specifically, proportionate to the retributivist desert an offender possesses.

So how is Lippke's negative retributivism *not* positive retributivism where crime reduction plays no part? Lippke admits his understanding of negative retribution is 'a more retributively-flavored theory of legal punishment' than it is often believed to be.[31] While acknowledging that there might be some exceptional circumstances where individuals are found to be so dangerous that their imprisonment beyond their original sentence might be warranted on some views of negative retributivism, it is unclear on what grounds this would be true for Lippke especially where he appears not to accept this as a problem for his own view.[32]

The only comment about non-desert factors playing some role in his theory arises in his discussion about how punishment as a practice ought not to degrade those punished. Lippke states that

[29] Brooks, *Punishment*, p. 99.
[30] Lippke, "Elaborating Negative Retributivism", p. 58.
[31] Lippke, "Elaborating Negative Retributivism", p. 60.
[32] Lippke, "Elaborating Negative Retributivism", pp. 61-62.

this 'non-degradation constraint' is 'like the more familiar retributive constraints' and so does appear to exclusive to negative retributivism and not available to positive retributivism.[33] He says: 'Put simply, we will see less crime in the future if offenders are not degraded (as the retributive constraint enjoins) but also prodded and helped to be morally responsible'.[34] In other words, if we punish offenders who are deserving and to the degree deserved, we should recognise that our imposition of punishment should attempt to enable offender rehabilitation by not degrading prisoners and developing their sense of moral responsibility. Rather than elaborating negative retributivism, Lippke appears to defend a position similar to positive retributivism. He avoids the problem of theoretical incoherence I highlighted with negative retributivist accounts by marginalising any role played by crime reduction. Note that the reason we should not punish disproportionately—either too much or too little than deserved within a range—is because of concerns that it might damage an offender's sense of moral responsibility. Note further that the reason we should not degrade offenders is because of the same concern. An offender's lack of moral responsibility is not simply a failure to rehabilitate and risk of reoffending, but primarily a failure to take sufficiently seriously the link between desert and punishment. However, it is claimed a retributivist justification and imposition of punishment should contribute to less criminal offending because there should be sufficient importance placed on developing an offender's moral responsibility.

Let me highlight this important point before turning to other concerns. Lippke convinces me here and elsewhere on many points in legal theory—and chiefly on how our theories of

[33] Lippke, "Elaborating Negative Retributivism", p. 67.
[34] Lippke, "Elaborating Negative Retributivism", p. 63-64.

punishment too often fail to account for their relation to practices. Lippke and I may disagree on how much of a negatively retributivist view he presents here, but I accept that any retributivist theory of punishment ought to share the concerns about an offender's moral responsibility raised first by him.[35]

There are two striking features of Lippke's account not already touched on. Note Lippke's claim that punishment should help to make offenders 'more morally responsible.'[36] This position appears to echo the claim that punishment should be rehabilitative through some form of moral education. The best exponent of this view is Jean Hampton:

Thus, according to moral education theory, punishment is not intended as a way of conditioning a human being to do what society wants her to do (in the way that an animal is conditioned by an electrified fence to stay within a pasture); rather, the theory maintains that punishment is intended as a way of teaching the wrongdoer that the action she did (or wants to do) is forbidden because it is morally wrong and should not be done for that reason.[37]

Both Lippke and Hampton appear to share the view that punishment should aim to make offenders more morally responsible. If successful, then offenders will refrain from future offending. Through educating offenders about their *criminal* wrongs as a kind (or kinds) of *moral* wrongs, we can reduce crimes by improving moral responsibility and awareness.

This view rests on an important mistake highlighted by my discussion in *Punishment*. The mistake is that not all crimes are

35 See Brooks, *Punishment*, pp. 225, fn. 3.
36 Lippke, "Elaborating Negative Retributivism", pp. 63-64.
37 Jean Hampton, "The Moral Education Theory of Punishment", *Philosophy and Public Affairs* 13 (1984): p. 212.

immoral and not all immorality is criminal. There is a '*justice gap*' too often overlooked between where moral education might be a relevant possibility and those crimes for which it is not.[38] This gap speaks to the distinction of *mala in se* crimes and *mala prohibita* crimes. The former are thought wrongs independent of their criminalisation by law; the latter are thought wrongs because of their criminalisation. Crimes commonly understood as kinds of *mala in se* are murder and theft. *Mala prohibita* crimes may include drug and traffic offences as well as prostitution although this category is more controversial. My first point is that if there is such a distinction to be made then it is clear not all crimes are moral wrongs and so Lippke's (and Hampton's) aim to rehabilitate through heightened moral sensibility might be irrelevant or fall short.

But even if we reject there are *mala prohibita* crimes, then it remains true that most offences included in the criminal law are *strict liability* offences where culpability is irrelevant. The bare fact that someone drove a car on a street above a speed limit is necessary and sufficient to justify a conviction for a traffic offence—and excessive speeding can lead to imprisonment lest this be seen as a trivial illustration. My point is that if not all *criminal* wrongs are *moral* wrongs, then moral education aimed at raising sufficient awareness of an offender's moral wrongdoing in offending misses its target. For Lippke, 'we will see less crime in future', in part, if offenders are 'helped to be more morally responsible' (5). But if the issue is instead *legal* responsibility (and not *moral* responsibility), such a crime reduction effort may underperform or even ineffective.

Now let us turn to Lippke's discussion of my unified theory of punishment. While we agree on the important link between rights

[38] See Brooks, *Punishment*, p. 57.

and punishment, there are issues worth clarifying further. First, he claims that I am 'on the right track in pointing to a theory of human rights and the protection of such rights within a legal scheme as providing some of the conceptual and normative backdrop for a theory of legal punishment' (7).

This mistakes my use of *rights* for *human rights*. I understand these differently whereby human rights—from my explicitly non-natural law perspective—are inclusive of those human rights found in international agreements, such as the European Convention on Human Rights or the UN's Universal Declaration of Human Rights. *Rights* are different and represent a community's recognition of freedoms worthy of protection, and may include a special acknowledgement of human rights. I argue that 'the criminal law aims at the protection of individual *legal rights*. Our legal rights are substantial freedoms worthy of protection for each member'.[39] I further clarify my views on the relation between freedom and rights by claiming it is 'broadly consistent with some versions of the capabilities approach, but note that the view of freedom used here may be consistent with several different theories of freedom'.[40]

This is a key point because it makes clear that the kind of *rights* I am discussion are not *human rights per se*. One reason would be that it is unclear that every part of the criminal law we might want to include in our criminal law is concerned with *human* rights alone (that may have a more universal character) than individual legal rights (that might differ from one political community to the next). It is clear that we have rights of movement that can pertain to any defensible view of traffic offences, but it is far from clear how they relate to human rights any better.

[39] Brooks, *Punishment*, p. 127.
[40] Brooks, *Punishment*, p. 236, fn. 21.

This point matters because my unified theory links the proportionality of punishment to the centrality of the right infringed or threatened by a crime. Lippke claims I run with three different possible meanings of what a restoration of rights might entail. The first is about *any* rights, such as to restitution and including conduct addressed by private law.[41] While it is true that rights are protected by more areas of law than the criminal law alone, my focus is clearly on the criminal only. Issues about contract and tort law are interesting, but not part of my examination of punishment and its justification. The second possible meaning Lippke claims to find is a 'censuring aspect' whereby punishment has some expressivist function.[42] As should now be clear, I do not deny that punishment can be understood—at least metaphorically—as an expression of public censure, but my view rejects expressivist theories Finally, Lippke claims my discussion of restoration also appears to support the view that punishment aims to reassure the public that rights shall be protected and laws reliably enforced. This is broadly more accurate of my view than the first two which I'd reject. But Lippke then raises the concern that punishment 'curtails or infringes the rights of offenders' and so seems counterproductive as a project of rights protection.[43] My argument is that through the use of punishment it can be possible to best maintain and protect our rights. Limiting another's freedom by requiring treatment for serious conditions that have contributed to persistent reoffending is a means to the maintenance and protection of rights not only for the rest of us should reoffending be reduced, if not stopped, but also for the offender. Lippke's criticism would have greater force if punishment was an end in itself. If we punished for its own sake, then it is clearer how

41 Lippke, "Elaborating Negative Retributivism", p. 69.
42 Ibid.
43 Ibid., p. 70.

restricting rights can pose problems. But if we punish as a means to another good like securing rights, then restricting rights might be justified as a measure of last resort where there is no better alternative to protecting and maintaining rights. And as it should be.

IV

Conclusion

I am especially grateful to Duff, Dempsey and Lippke for these thoughtful and largely constructive comments on *Punishment*. While I can't say that I am convinced my views on punishment should change, these critiques provide a welcome opportunity to spell out in further detail the reasons behind the arguments I offer. I hope they may even shed some further light.

In conclusion, I would like to comment further on two points that arose during a conference organised by the editors of *Philosophy and Public Issues* held at LUISS this past spring. The first point is I was pushed to say more about why punishment should be unified. On the one hand, I appear to align theory to practice. I note that the Model Penal Code and sentencing guidelines across multiple jurisdictions include multiple penal purposes, but without a satisfactory framework for resolving any conflicts between these purposes when applied in practice. So is the unified theory about justifying our practices? This would seem to fit with my broadly Hegel-inspired work, as Hegel saw his philosophy as an effort at discerning the rationality in the word.[44]

[44] See Thom Brooks (ed.), *Hegel's Philosophy of Right* (Oxford: Blackwell, 2012) and Thom Brooks, *Hegel's Political Philosophy: A Systematic Reading of the Philosophy of Right*, 2d (Edinburgh: Edinburgh University Press, 2013).

Am I doing the same? On the other hand, I appear to be trying to provide a coherent theory about how a unified theory of punishment is possible. So is my aim to provide a theory of punishment or to justify our existing practices?

The short answer is a bit of both. My view is that a coherent, unified theory of punishment is possible and part of its wider importance is it can offer us a possible framework to guide existing sentencing policy. But it is not the bare existence of these policies that provides my primary philosophical motivations, but they are also not irrelevant. A unified theory is not only possible, but it also highlights a neglected tradition of Hegelian thought so there is some importance for the history of ideas from my theory of punishment as well.[45] But I do not assume our practices are correct or desirable. We should not be interested in a unified theory because our practices cover plural purposes, but instead because these practices get right that these purposes are worth having for sentencing—so what we require is a new framework which my unified theory attempts to provide.

A second point concerns the movement of travel. I focus on rights to be protected and move from there. But it might be objected that I should start with wrongs and go to rights. The problem is that I run a risk of resting my view on an overinflated

[45] See Thom Brooks, "Is Hegel a Retributivist?", *Bulletin of the Hegel Society of Great Britain* 49/50 (2004): pp. 113-26; Thom Brooks, "Rethinking Punishment", *International Journal of Jurisprudence and Philosophy of Law* 1 (2007): pp. 27-34; Thom Brooks, "Punishment and British Idealism", in Jesper Ryberg and J. Angelo Corlett (eds.), *Punishment and Ethics: New Perspectives* (Basingstoke: Palgrave Macmillan, 2010): pp. 16-32; Thom Brooks, "Punishment: Political, Not Moral", *New Criminal Law Review* 14 (2011): 427-38; Thom Brooks, "Is Bradley a Retributivist?", *History of Political Thought* 32 (2011): pp. 83-96 and Thom Brooks, "Hegel and the Unified Theory of Punishment", in *Hegel's Philosophy of Right*, pp. 103-23.

view of rights.[46] While I accept that this risk is a concern, I remain unconvinced the alternative mentioned would better avoid this problem.

A book is more than a series of claims and arguments. I spent several years researching, constructing and rewriting the text to cover necessary ground and clarify my positions. After such a major effort, it is immensely satisfying to receive such robust and wide-ranging commentary from so many philosophers I highly respect. I hope these comments go some way to pay back this kindness.[47]

Durham University

[46] I am especially grateful to Vittorio Bufacchi for raising this concern.

[47] I am very grateful to Gianfranco Pellegrino, Michele Bocchiola, Vittorio Bufacchi, Michele Mangini and Mario Ricciardi for comments and discussion on *Punishment* during my visit to LUISS earlier this year.

If you need to cite this article, please use the following format:

Brooks, Thoms, "Defending Punishment. Replies To Critics," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), 73-94, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

THE PHILOSOPHY OF PUNISHMENT

PUNISHMENT WITHOUT PAIN

OUTLINE FOR A NON-AFFLICTIVE DEFINITION OF LEGAL PUNISHMENT

BY

ANDREI POAMA

© 2015 – *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015): 97-134
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Punishment Without Pain
Outline for a Non-Afflictive Definition of Legal Punishment

Andrei Poama

The intentional imposition of suffering on offenders is typically taken to be one of the necessary components of the *definiens* of legal punishment. Call this the afflictive definition of punishment. I argue that, despite its intellectual pedigree and contemporary pervasiveness, the afflictive definition is flawed in at least three ways: it is ambiguous, inaccurate, and non-operationalizable. In order to escape these three problems, I suggest that we should amend the definition of punishment by taking *suffering* out of its *definiens* and replacing it with the idea of *sanctions*. Call this second, amended definition the non-afflictive definition of punishment. I believe that my case is strongest in arguing for the elimination of suffering from our definition of punishment, and that my proposal provides a plausible solution for a sounder definition of legal punishment.[1]

The argument is structured as follows. In Section I, I offer some background for my critique, by tracing it back to Jacob Adler's denunciation of what he calls the standard view of

[1] In what follows, I am using the terms 'definition', 'construal' and 'understanding' in an interchangeable way, unless otherwise explicitly stated.

Philosophy and Public Issues (New Series), Vol. 5, No. 1 (2015): 97-134
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

punishment.[2] In Section II, I formulate my threefold critique of the afflictive definition of punishment. I argue that, whatever one's meta-theory of definition is, there are good reasons to require that any adequate definition of punishment be reasonably clear when it comes to understanding what punishment is (*non-ambiguity*), sufficiently accurate so that it does not exclude typical penal sanctions or include non-penal acts (*accuracy*), and suitably formulated to allow us to decide whether we are actually engaged in imposing punishment on someone or not (*operationalizability*).[3] The afflictive definition fails to meet each of these three minimal definitional criteria. In Section III, I suggest that we should replace the afflictive definition with a non-afflictive one, and that we should do this by eliminating the idea of suffering in favor of that of sanctions. I also explore two possible objections to this definitional reform: the objection from circularity and the objection from concealment. Moreover, I single out two implications of disposing of the afflictive definition. The first implication is that a non-afflictive definition proves neutral as seen from the standpoint of various potential justifications of punishment. The second one is that a non-afflictive definition of punishment allows us to see that penal abolitionism rests on a definitional mistake.

[2] Jacob Adler, *The Urgings of Conscience. A Theory of Punishment* (Philadelphia: Temple University Press, 1992), 83-90.

[3] If your meta-theory of definition is a sceptical one, and holds that any definitional attempt is a non-starter, then my whole argument will fail to convince you. That should not bother me too much, since I am concerned with an audience that takes definitions seriously. But, if you think that a minimal definition of any (normative) concept (or term or thing) is in order, then you should take the three criteria that I am putting forward to work independently from any particular understanding of what an appropriate definition should amount to substantively (in terms of the *nature* of the definitional activity) or methodologically (in terms of the *way* in which a definition should be formulated).

Before moving on to the first section, three preliminary remarks are in order. The first one concerns the status of the proposal for a non-afflictive definition of punishment. Though it might seem trivial to some, it is important to underline that the definition that I am aiming for is not a *legal* definition of legal punishment, but a theoretical definition of legal punishment. A legal definition of punishment is literally missing from our penal codes or otherwise legally authorative texts.[4] The main point worth noting at this level is that a strictly theoretical treatment of the question of punishment could never count as a legal one. For a definition to count as legal, we would have to act in a legislative setting and be vested with the appropriate competence for devising authoritative definitions. This remark is important insofar as it makes clear why an adequate definition of punishment should not be legislative in form or intent: we do not get to stipulate our way into the meaning of punishment. Consequently, this article should not be read as a proposal for a new understanding of punishment, but as a report on its current legal meaning.

The second remark is that this is an analysis that deals with legal punishment in particular, not with punishment in general. Addressing the question of punishment *simpliciter* is significant in its own right, but it tends to downplay the differences between the content of various forms of punishment, the identity of the agents that impose or experience them, and the limits set on the severity of those sanctions. Grounding a child is not punishing in

[4] This is the case for the U.S. *Model Penal Code*, the French *Code Pénal*, the Belgian *Code Pénal*, the German *Strafgesetzbuch*, but the list is longer than that. Also, it is worth noticing that the absence of a legal definition is not limited to the question of punishment. Other terms—such as 'income'—do not have a clear-cut legal definition either. For a more detailed discussion, see Huntington Cairns, "A Note on Legal Definitions," *Columbia Law Review* 36, no. 7 (1936): 1099-1106.

the same sense as imprisoning an offender is, and any definition of punishment *simpliciter* will unavoidably miss the difference between these two kinds of penal practices. One specifying feature of legal punishment that will get lost in any analysis of punishment *simpliciter* is the fact that the former is bound by the principle of legality. One could hardly talk about forms of legal punishment that could not minimally align with this principle.[5] However, since legal punishment is a particular instance of punishment in general, any definition of the former will have an important bearing on the definition of the latter. If suffering should be taken out of the *definiens* of legal punishment, then it should be equally eliminated from that of punishment *simpliciter*.[6]

The third, and final, remark is that I am concerned with the definition of punishment, not with its justification. Definition and justification are obviously linked, but they should be kept distinct. One of the reasons for doing so is that we do not want to gerrymander our definition of punishment according to our preferred justification. Another reason is that definitional matters should be satisfactorily settled prior to normative debates, insofar as an unstable definition of punishment will be vulnerable to *ad-hoc* reinterpretations in a way that serves our privileged

[5] I take the legality principle to refer to the requirement according to which there is no punishment in the legal sense of the term unless there is a previously law specifying both the offence to be punished and its corresponding punishment. Thus read, the legality principle excludes the possibility of *legally* punishing someone retroactively, that is, for an act that was not a legal offence at the time when it was committed. This allows, of course, for the conceptual possibility of punishing that same act in a non-legal way. See Gabriel Hallevy, *A Modern Treatise on the Principle of Legality in Criminal Law* (Berlin, Heidelberg: Springer-Verlag, 2010).

[6] For a definitional analysis of punishment *simpliciter*, see Leo Zaibert, *Punishment and Retribution* (Aldershot: Ashgate Publishing, 2006).

justificatory account of punishment.[7] The priority of definitions that I have in mind is of a logical kind. Whenever we disagree about the adequate justification of punishment, we should be able to assess whether we are referring to the same thing or concept. Otherwise, we risk talking past each other. For example, though Rawlsian-minded liberals and communitarians disagree about the appropriate principles of distributive justice, their debate remains intelligible only insofar as they agree that distributive justice should be defined as the allocation of certain goods among the members of a society. The content and scope of what counts as a *good*, as a *member* or even as a *society* remain open to normative disagreements. But these disagreements make sense only against the background of a commonly held definition of justice.[8]

I

Shifting Away from Suffering: Background of the Non-Afflictive Definition of Punishment

Defining punishment without resorting to the idea of suffering is not an original undertaking, but it remains an exceptional one.

[7] For this last second point, see Antony Flew, "The Justification of Punishment," *Philosophy* 29, no. 111 (1954): 291-307.

[8] The logical priority of definitions over justifications does not exclude the possibility of disagreements about definitions. Moral, political, and legal philosophy is replete with both definitional and normative debates. My argument is simply that a theory cannot be said to win a disagreement about the justificatory principles of a practice by surreptitiously resorting to a different definition of its subject-matter and thus changing the terms of the debate. It is nonetheless true that normative disagreements might sometimes lead to definitional debates. This happens when supporters of a specific theory argue that their opponents do not properly understand the meaning of a specific normative practice or concept. My claim, then, is that, if theorists radically disagree over the definition of their subject-matter, their *normative* disagreement will most likely never get off the ground.

A large majority of legal theorists and penal philosophers take suffering to be a necessary feature of punishment. Only a few of them construe punishment in the absence of any reference to afflictive experience, be it suffering, pain, unpleasantness, harm, evil or other sub-varieties of affliction.[9] Even fewer authors explicitly go against understanding punishment in terms of suffering. Jacob Adler is, in this sense, an exception.[10] Adler criticizes the conceptions according to which punishment inherently involves the experience of suffering by offenders. He takes these conceptions to form what he calls the standard view of punishment, an outlook that he deems representative of the ways in which punishment has been traditionally portrayed in the history of Western legal and political thought, from Plato, through Aquinas, Hobbes, Kant and up to contemporary figures like Hart or Nozick.[11]

[9] Some of these rare authors—namely, John Rawls, Claudia Card, Hugo Bedau—are enumerated and referenced in Adler, *The Urgings*, 80-108.

[10] For another exception (quoted by Adler), see Unto Tähtinen, *Non-violent Theories of Punishment: Indian and Western* (Motilal Banarsidass, 1983). For a recent attempt at dissociating between burdensomeness and the intentional imposition of suffering, see Bill Wringe, "Must Punishment Be Intended to Cause Suffering?," *Ethical Theory and Moral Practice* 16 (2013): 863-877. Wringe's project differs from the one that I pursue in this article. I am strictly concerned with definitions, whereas Wringe examines both definitional and justificatory questions. For yet another critique of the afflictive definition of punishment, see Helen Anne Brown Coverdale, *Punishing with Care: treating offenders as equal persons in criminal punishment* (LSE PhD Thesis, October 2013), available at: etheses.lse.ac.uk/1080/.

[11] For a complete list of references, see Adler, *The Urgings*, idem. For other afflictive-prone authors that Adler might have taken into account but did not, see James Smith, "Punishment: A Conceptual Map and a Normative Claim," *Ethics* 75, no. 4 (1965): 285-290; Sidney Gendin, "The Meaning of 'Punishment'," *Philosophy and Phenomenological Research* 28, no. 2 (1967): 235-240; C.J. Ducasse, "Philosophy and Wisdom in Punishment and Reward," IN *Philosophical Perspectives on Punishment*, eds. Edward H. Madden, Rollo Handy and

Another way of formulating the standard view of punishment—which roughly corresponds to the afflictive definition—is, by way of metonymy, to speak about the Flew-Hart-Benn definition.[12] This latter definition lists five independently necessary and jointly sufficient conditions (or criteria) for an action or practice to count as punishment. According to this definition, punishment should consist in (1) the intentional inflicting of suffering, (2) on putative offenders, (3) for the offences that they are judged to have committed, (4) by someone else than the offender, and (5) holding a special authority qualified in terms of specific institutional rules.

Like Adler, I claim that the notion of suffering should be abandoned when it comes to understanding punishment, though my reasons are not the same. Adler's rejection of the standard view rests on two arguments. The first argument is that there are important counter-examples to suffering-centred forms of punishment, such as legal community service sanctions, push-ups in the case of cursory military sanctions or penance rituals pertaining to religious penal practices.[13] The second argument is that suffering should be eliminated from our ideal accounts of justified instances of punishment. Because it explicitly rests on a

Marvin Farber (Springfield, IL: Charles Thomas, 1968), 3-19; Joseph Margolis, "Punishment," *Social Theory and Practice* 2, no. 3 (1973): 347-363; Walter Kaufmann, *Without Guilt and Justice. From Decidophobia to Autonomy* (NY: Peter H. Wyden Inc., 1973); J.P. Day, "Retributive Punishment," *Mind* 87 (1978): 498-516; J.R. Lucas, *On Justice* (Oxford: Clarendon Press, 1980); Philip Bean, *Punishment: A Philosophical and Criminological Inquiry* (Oxford: Martin Robertson, 1981); Kent Greenwalt, "Punishment," *Journal of Criminal Law and Criminology* 74, no. 2: 343-362; Steven Sverdlik, "Punishment," *Law and Philosophy* 7, no. 2 (1988): 179-201.

12 Flew, "The Justification"; H.L.A. Hart, *Punishment and Responsibility* (Oxford: Clarendon Press, 1968), 8-11; S.I. Benn, "An Approach to the Problems of Punishment," *Philosophy* 33, no. 127 (1958): 325-341.

13 Adler, *The Urgings*, 91-100.

'claim about which cases of punishment are important or ideal,' Adler's critique of penal suffering is a normative one. His standpoint is justificatory, not definitional.[14]

Both arguments are inadequate for the purpose of defining legal punishment. The limits of the first argument are obvious. The examples of military or religious sanctions do not have any decisive bearing on the way in which we should construe state-enforced legal punishment. More generally, such an argument does not exclude the possibility of multiplying the list of counter-examples *ad libitum*—something that Adler does when he mentions the eccentric counterfactual examples of one-minute long prison sentences or ten cents fines—in a manner that would weaken the accuracy and comprehensiveness of any definitional feature of punishment.[15]

The second argument is faulty insofar as it fuses an ideal construction of punishment with its definition. Adler's aim in defending a suffering-free interpretation of punishment is to set the ground for a justification of penal practice that could not hold if punishment were to be intrinsically characterized by suffering, harm, evil, pain or other forms of affliction.[16] This makes his critique of penal suffering seem *ad-hoc* and non-neutral. We cannot—or, at least, should not—arrange for the appropriateness of our justificatory accounts by tweaking the meaning of the practices we want to justify until they align with our normative commitments. If such moves were allowed, then any justification would be possible, granted that we have stipulated what the ideal

[14] Id., 81.
[15] Ibid.
[16] Adler holds that punishment is justified by a principle of rectification of offences, which are in turn construed as violations of basic rights.

meaning of a given practice should be for the purpose of *our* desired normative standpoint.[17]

The limits of Adler's two arguments might explain why his critique of penal suffering has not been taken seriously enough. Legal and political philosophers today continue to define punishment in terms of suffering. For example, Claire Finkelstein insists that punishment inherently 'involves the infliction of pain or other form of unpleasant treatment.'[18] Antony Duff takes punishment to be 'something intended to be burdensome or painful.'[19] Daniel McDermott considers that the fact that 'punishment *must* cause suffering' makes it a stringent subject-matter for moral consideration.[20] Steven Tudor thinks that it is 'uncontroversial that punishment, by definition, involves suffering.'[21] Leo Zaibert asserts that punishment should be understood as 'the general phenomenon whereby we inflict something we believe is painful for the wrongdoer as a result of her wrongdoing.'[22] Mitchell Berman argues that punishment

[17] My claim here is that Adler is right about legal punishment not being definitionally afflictive, but that defending a non-afflictive definition on the basis of one's particular normative conception of punishment is not the right kind of reason from a strictly definitional standpoint. In other words, one is not entitled to resort to a particular justification of punishment to show why its afflictive definition is unwarranted.

[18] Claire Finkelstein, "Positivism and the Notion of an Offence," *California Law Review* 88: 358.

[19] Antony Duff, *Punishment, Communication, and Community* (Oxford: Oxford University Press, 2001), XIV-XVI.

[20] Daniel McDermott, "The Permissibility of Punishment," *Law and Philosophy* 20, no. 4 (2001): 403.

[21] Steven Tudor, "Accepting One's Punishment as Meaningful Suffering," *Law and Philosophy* 20, no. 6 (2001): 583.

[22] Zaibert, *Punishment and Retribution*, 36. As already indicated, Zaibert examines punishment *simpliciter*, but his definition applies, by way of logical consequence, to legal punishment as well.

presupposes the 'imposition of something painful or burdensome.'[23] Daniel Boonin affirms that 'intent of harming' is a necessary definitional feature of punishment.[24] Nathan Hanna similarly claims that punishment invariably 'inflicts pain, suffering or burdens.'[25]

The examples could be multiplied, but I take it that these recent illustrations, Adler's initial list and the complementary references that I enumerate in footnote 10 are sufficient to show that there is a strong consensus that punishment involves suffering as a matter of definition. One could nonetheless notice that the term *suffering* does not appear in all the definitional formulae. I am not worried about this terminological variation. This is because whatever I find problematic about penal suffering will also apply to penal pain, unpleasant treatment, burdensomeness, harm or to whatever other expression used to convey the idea of penal affliction.[26] Another reason not to care

[23] Mitchell L. Berman, "Punishment and Justification," *Ethics* 118, no. 2 (2008): 261.

[24] David Boonin, *The Problem of Punishment* (Cambridge: Cambridge University Press, 2008), 25.

[25] Nathan Hanna, "Liberalism and the General Justifiability of Punishment", *Philosophical Studies* 145, no. 3 (2009): 329.

[26] One could argue that the burdensome character of legal punishment is not the same as its afflictive character, insofar as burdensomeness does not necessarily entail suffering. If this is correct, then my critique does not apply to those authors who define punishment in terms of the intentional imposition of burdensomeness. However, even those authors who use the terms *burdensome* and *burden* in addition to other terms like *pain*, *suffering*, *unpleasant treatment* or *harm* do not insist on the categorically different meaning of the former as compared to the latter. One can then reasonably assume that they use these terms interchangeably and not disjunctively. Moreover, if *burdensomeness* understood non-afflictively were a necessarily defining feature of legal punishment, but painfulness, suffering or harmfulness were not, it is not clear why the authors I have cited choose to include both the former and the latter in the *definiens* of punishment.

about terminological variation is that the vocabulary of suffering seems to be more widely used than others. The language of suffering could, from this point of view, be representative of what is generally meant when philosophers define punishment by resorting to the language of affliction.

The goal of this section was twofold. First, I wanted to make clear that it is not for the first time that an afflictive understanding of punishment is being denounced. Second, I tried to show that the definition of punishment in terms of suffering is still the rule—not to say the habit—among legal and political philosophers. One possible explanation as to why this is so is that a convincing critique of the afflictive definition *qua* definition is missing. I try to offer such a critique in the following section.

II

The Afflictive Definition of Legal Punishment: Three Flaws

In criticizing the afflictive definition, I do not intend to address any meta-theoretical debate about the right purpose of definitions or about the adequacy of different definitional methodologies.[27] I have two reasons for not doing so. First, turning to a debate about the definition of *definition* is bound to take us to considerations whose contested character and generality will not contribute to our understanding of legal punishment in any interesting way. Second, my critique of the afflictive definition does not depend on any prior commitment to a particular view about the nature of definitions or the correct definitional methodology. In other words, I think that the grounds for going against the afflictive definition cut across our

[27] For a systematic discussion of the relevant meta-definitional debates, see Richard Robinson, *Definition* (Oxford: Oxford University Press, 1963).

meta-theoretical disagreements about definitions in general. Put differently, my proposal for eliminating the afflictive definition does not depend on any narrow or contested view of what a definition is or does.

The three objections that I raise against construing punishment in terms of suffering rely on three definitional criteria. These three criteria are non-ambiguity, accuracy and operationalizability. Before moving to the crux of my critique, I want to make explicit what I understand by each of these terms. Since their content is quite straightforward, not much explaining will be actually needed.

I start with non-ambiguity, which I take to refer to the absence of uncertainty or unmanageably multiple meanings of the *definiens*. Non-ambiguity is not an absolute criterion, since the definition of a term will not suffice to guarantee its disambiguation in other respects. There is no straight line leading from a non-ambiguous definition to a non-ambiguous use of a particular sign or concept. As Robinson puts it, 'we should always have in mind the probability of ambiguity and the flexible nature of our vocabulary which causes it.'[28] This is certainly something to be aware of. However, the fact that ambiguity is not totally eliminable does not mean that it should not be minimized or made explicit if we want to avoid its morphing into equivocation.

We should be particularly interested in coming up with a non-ambiguous definition of legal punishment. This is because an ambiguous understanding of punishment could not accommodate the legality requirement and would thus fail to be a definition of legal punishment proper. If the *definiens* of legal punishment rests on multiple meanings, this should be clearly stated. But multiplicity of meanings would be awkward in this case. When

[28] Id.,154.

asked what we understand by legal punishment, we do not envisage a plurality of equally valid meanings and then choose one meaning in particular, even though we could have just as well chosen a different one. The task of a non-ambiguous definition of legal punishment is to provide us with a procedure for assessing whether we are actually talking about the same thing or concept, even if—and especially when—our interpretations or justifications of punishment differ.

The second, accuracy criterion states that a definition of legal punishment should be constructed in a way that is extensionally appropriate. This means that the definition should be sufficiently wide to include all instances of legal punishment and narrow enough to exclude instances that do not fall under our understanding of it. In particular, we should consider that a definition is radically inaccurate if it fails to include paradigmatic cases of legal punishment. A definition of punishment will be moderately inaccurate if it cannot accommodate instances of punishment that are not necessarily typical of penal practice. I will argue that the afflictive definition is both moderately and radically inaccurate.

The third, operationalizability criterion requires that the definition of legal punishment serve as a suitable basis for developing an operational definition of it. This is not the same as demanding that the definition of punishment *be* an operational one. Rather, the idea is that any adequate definition of punishment should offer us a good insight into the way in which we can go about formulating a subsequent operational definition. This is not an idiosyncratic requirement. Even in the case of the theoretical definitions of physical objects, we take it to be an advantage of these definitions that they can support us into identifying adequate procedures for constructing or representing those same objects. For example, defining weight as the force

exerted on an object due to gravity guides us into operationally assessing the presence and value of weight as the result of the measurement of objects on a Newton spring scale. When it comes to legal punishment, an operational definition is important because it provides us with a procedure for testing whether we are actually engaged in penal practice or not. More generally, operationalizability is what makes a definition practically relevant. As I will try to show, the afflictive definition of punishment is practically irrelevant because it is non-operationalizable.

The Problem of Ambiguity

I now turn to my first objection, according to which defining punishment in terms of the intentional infliction of suffering is ambiguous. There are two ambiguity problems in the afflictive definition. The first problem concerns the meaning of the term 'intentional', whereas the second one pertains to the multiple or sometimes uncertain meanings of the term 'suffering.' Since it can be easily bypassed, we can call the first problem the easy ambiguity problem. Because I do not see any convincing method for avoiding the second problem, I suggest that we call it the hard ambiguity problem.

The easy ambiguity problem is that the term 'intentional' is ambiguous between motives and objectives. It is unclear whether what is meant by saying that punishment is the intentional infliction of suffering should be construed as referring to the subjective motives of the penal agents or to the objective aims of the penal institution. However, we have good reasons to think that penal intentions apply to objectives and not to motives. If intentions were about motives, punishment would become quite an erratic practice. It is very difficult to test whether the representatives of the penal institution are actually motivated by

the offenders' suffering. Moreover, interpreting intentions in terms of motives would render punishment dependent on the actual existence of persons with a disposition for imposing suffering on other persons. Worse, any proof that penal agents lack afflictive motivation would authorize us into looking for agents that have such a motivation and solicit them to visit real punishment on offenders. This would make punishment a highly unpredictable practice and would violate the legality requirement.

Fortunately, we can resolve the easy ambiguity problem by specifying that intentions refer to objectives and not to motives. Unfortunately, this does not help us to solve the hard ambiguity problem, which is related to the multiplicity and uncertainty of the meaning(s) of suffering. It is not clear what the content of suffering is or should be. We can single out at least four dimensions that are constitutive of its ambiguity.[29] The first dimension concerns the variation in different kinds of suffering. There is, first, physical suffering, such as physical pain, discomfort or exhaustion. Second, suffering can express itself in a psychological mode, if we consider phenomena like fear, depression, shame, humiliation, anxiety, panic, and so on. Third, one should take into account more existential or moral forms of suffering that do not have to be caused by an external stimulus, but which can be enabled by a particular context or the performance of a specific action. This is the case of remorse, repentance, regret, grief and other similar experiences of contrition.[30]

[29] The ambiguity generated by different kinds of suffering is explored in Jamie Mayerfeld, *Suffering and Moral Responsibility* (Oxford: Oxford University Press, 2002).

[30] Arguing that this dimension of ambiguity can be avoided by stipulating that penally relevant suffering should be restricted to a single kind—say, physical suffering—would be an arbitrary move. This is mainly because saying that

The second dimension of ambiguity is closely related to the first one. It emphasizes the difficulty of specifying the structure or content of the combination of different kinds of suffering that one should pursue when engaged in the practice of punishment. Combining different kinds of suffering into a suffering function might be impossible if we realize that these different kinds of suffering are incommensurable. One cannot try to combine entities that are categorically different, just as one cannot add numbers and letters if the latter are considered *qua* letters and not as algebraic expressions. Similarly, it does not seem to make sense to say that physical exhaustion is intra-personally commensurable with social humiliation or with moral regret.

The third dimension of ambiguity is not linked to our difficulty in constructing appropriate intra-personal comparisons between different kinds of suffering, but to the predicaments concerning the inter-personal comparisons between different subjective experiences of suffering. The punishment of some offenders might give rise to a strong physical sensation of discomfort, whereas the punishment of others could bring about an intense impression of social discrimination or stigmatization. If we take the afflictive definition as a guide into understanding the meaning of punishment, it is not obvious whether a penal sanction that results in physical discomfort and one that results in

punishment that does not track our stipulated form of suffering is not punishment in *our* sense of the word can always exclude other positively existing forms of legal punishment. For example, if our stipulation restricts suffering to the physical kind, it would be difficult to see how a fine or community work might be singled out as punishments from the standpoint of our stipulative definition. As argued on page 2, the stipulative move is open to the legislator, but it is not a theoretically sound move.

a sense of discrimination should be construed as comparable cases of punishment in any minimally informative way.[31]

The fourth dimension of ambiguity derives from the multiple ways in which penal suffering can be temporally organized. It is not clear *when* suffering should occur if we are to envisage it in terms of penal suffering proper. It is quite clear that suffering experienced by someone who experiences anxiety or fear at the mere thought of future punishment is not, properly speaking, a part of punishment, because any person experiencing such an anxiety or fear could then claim that she is being punished. Such an outlandish claim can be reasonably dismissed and does not therefore raise a problem for the supporter of the afflictive definition.[32] Even so, it remains that the afflictive definition cannot help us decide whether the economic discomfort experienced upon an offender's attempt to regain a normal social life after release from prison should be read as penal suffering or not. Also, it is difficult to tell whether an offender was really punished if, for example, she experienced suffering only during the last moments of her prison term. Should we then say that the non-afflictive experience of imprisonment up to its last moments should not be understood as punishment?

The Problem of Inaccuracy

This fourth dimension of ambiguity leads me to consider my second objection against the afflictive definition. This is the

[31] For an article that highlights the non-fungibility of different forms of suffering, see David Gray, "Punishment as Suffering," *Vanderbilt Law Review* 63, no. 6: 1620-1693.

[32] Suffering experienced by someone at the thought of a probable punishment is also hard to consider as part of punishment proper, since it would collapse the positive distinction between the accused and the punished.

objection from inaccuracy. The objection can be formulated in two ways. First, one can say that an important number of penal sanctions are not very likely to produce suffering in any significant way. Sanctions like fines lack the affliction-prone dimension that might otherwise probabilistically characterize prison sentences. Second, one can come up with examples where offenders will either take pleasure in or be indifferent to the penal sanctions that are being imposed on them. Call the first inaccuracy problem the fine problem and the second inaccuracy problem the masochist/callous offender problem.

I will start with the fine problem. The reasoning behind this problem runs as follows: most penal sanctions do not, as a general rule, produce the kind of afflictive effects that would be required to take suffering as one of the necessary features of punishment. As it turns out, most penal sanctions are fines, and most of them are imposed for minor traffic offences.[33] It is

[33] This is at least the case for the U.S. and the French criminal justice statistics. For example, in France, the percentage varies between 30% and 41% in the 1990 to 2009 time period. See *Infostat Justice. Bulletin d'information statistique*, no. 114 (2011). The statistical reality of punishment matters, since we need a definition of real legal punishment, that is, a definition that has descriptive accuracy. It is also worth emphasizing that the legal qualification for minor traffic offenses in most U.S. state-level jurisdictions is penal, *not* civil. Out of the 52 state-level jurisdictions listed by the 2010 *Summary of State Speed Laws*, only 11 states—that is, 21%—have distinctly civil sanctions for minor speeding violations. The District of Columbia has both civil and penal sanctions for minor speeding violations. An overwhelming majority of the U.S. states have a penal qualification for minor speeding violations. The specific qualification varies from misdemeanours (Alabama, Arkansas, Delaware, Georgia, Indiana, Iowa, Louisiana, Maryland, Missouri, Mississippi, Montana, Nevada, North Carolina, Ohio, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming), to infractions (California, Colorado, Connecticut, Hawaii, Idaho, Kansas, Missouri, Nebraska, New York, North Carolina, Oregon, Virginia), violations (Kentucky, New Hampshire) or petty misdemeanours (Minnesota). See *Summary of State*

therefore reasonable to consider that most penal sanctions will not fit the afflictive definition, insofar as they will not result in the suffering of offenders.[34]

The supporter of the afflictive view of legal punishment might try to dismiss the fine problem by resorting to two counter-arguments. The first counter-argument is that the fine problem is not a problem at all, given that even the smallest fine will produce a level of dissatisfaction, discomfort or annoyance which can easily be translated in terms of suffering. This argument is defective in two respects. First, it tends to trivialize and dilute the idea of suffering. This is because it considers that the payment of a fine involves a kind of suffering which, though less intense, is not fundamentally different from the suffering incurred through other forms of punishment, such as imprisonment, community work or electronic surveillance.

The supporter of the afflictive definition might reply that fine-generated suffering is special. But if the suffering generated by a small fine is not of the same type—meaning that it does not have the same nature—as the one produced by imprisonment or other sanctions, then this first counter-argument rests on a fallacy of equivocation. One cannot attempt to define legal punishment *in general* as the intentional infliction of suffering and then claim that the kind of suffering comprised in certain forms of punishment is

Speed Laws, National Highway Traffic Safety Administration, 11th edition (2011). Traffic violations—whether illegal parking or speeding—are generally considered as belonging to the criminal law in the U.K. as well. For a more extensive analysis of traffic violations as criminal offences, see Thom Brooks, *Punishment* (Routledge, 2013), 20, or Sally Cunningham, *Driving Offences: Law, Policy, and Practice* (Aldershot: Ashgate, 2008).

[34] Saying that fines are usually ancillary to prison terms or to probation is not sufficient to show that fines are *themselves* afflictive, because we could still consider it possible for the pain of imprisonment to generate suffering even in the absence of an accompanying fine.

categorically distinct from the suffering contained in others. This fragments the idea of suffering in a way that renders the afflictive definition equivocal across different instances of punishment.

Second, arguing that even small fines necessarily or typically generate suffering tends to misrepresent what fining is all about, which is the payment of a sum of money. It is not very plausible to try to describe the payment itself in afflictive terms. Simply paying for something does not necessarily cause the experience of suffering. We give money for a lot of things, but we do not necessarily nor usually suffer *because* we do so. In other words, the reason for suffering does not naturally reside in the fact of paying. If this were the case, then paying for gifts, complying with taxation or reimbursing a loan would all have to result in some form of suffering. If they do not, then it is not immediately obvious that suffering is a necessary feature of payments. Without a doubt, the fact that the notion of 'suffering' and that of 'payment' are not definitionally linked does not imply that the offenders who are fined—and especially those who are economically disadvantaged—will never suffer. However, the *probable* suffering of financially impaired offenders is insufficient for concluding that suffering is a *necessary* feature of all payments and, consequently, of all fines.[35]

[35] I do not deny that financially impaired offenders might in some cases suffer more from paying a fine than from executing a prison sentence. Even so, four remarks are in order here. First, the perception of future suffering is not equivalent to the actual experience of suffering. Empirical evidence tends to show that people are generally bad at forecasting the extent to which they will be negatively affected by money losses, and that, given the phenomenon of hedonic adaptation, financial losses *do not* generate significantly or substantially higher levels of subjective disvalue. See, for example, John Bronsteen, Christopher Buccafosco, Jonathan S. Masur, "Happiness and Punishment," *The University of Chicago Law Review*, 76 (2009): 1037-1081. Second, in cases where financially impaired offenders cannot possibly pay their fines, they

The second counter-argument that might be advanced by the supporter of the afflictive definition is that fines do not matter, because they are not really punishments in the same way in which, for example, imprisonment or electronic surveillance are. This response seems to rely on a claim according to which the entire class of punishments should be construed according to a core-periphery model, with real punishments at the core and quasi-punishments at the periphery. If we accept this model, we could say that imprisonment should, in virtue of its severity or seriousness, be interpreted as a real form of punishment, whereas fines are more adequately understood as quasi-punishments.

However, the core-periphery model fails to support suffering as a necessary feature of punishment. This is because there is at least one form of punishment—namely, the death penalty—that will be intuitively situated at the core of the class of all punishments, without thereby necessarily causing the offenders' suffering.[36] This might sound like a strange claim, but it fits both

cannot be said to undergo fining. Third, a fine does not have to be paid in a single lump sum, but can be spread over a longer duration. Such an arrangement might lower the probability and impact of the respective fine-generated afflictive experience. Fourth, even if we agree that losing money entails some form of suffering, this loss is not in any way definitional of punishment. Other policies—for example, taxation—involve some form of financial loss as well. But it cannot be said that tax-caused financial losses entitle us to refer to taxation as being a form of punishment, at least if we agree on using the term *punishment* in a non-figurative way. More generally, I agree with Gray that suffering is an incidental or contingent effect of punishing, but not a necessary feature of the 'normative concept of punishment'. See Gray, *op. cit.*: 1623. I want to thank both reviewers for helping me to clarify what is exactly at stake in the fine problem.

[36] Only 98 out of 195 states in the world have legally abolished the death penalty for all criminal offences. Though more states now have a *de facto* moratorium on the death penalty, the legal execution of people is far from being a peripheral form of punishment. For a more detailed analysis of capital

with the history of the technological transformations in the practice of capital punishment and with the jurisprudence that specifies its meaning and content.

The history of capital punishment can be plausibly interpreted as a series of attempts to take physical suffering out of the process of execution. For example, as Pieter Spierenburg recalls, one of the main rationales for introducing the guillotine as an execution method in France at the end of the XVIII century is that it was considered to be a painless procedure.[37] The painlessness of decapitation was certainly contested, but only in the name of other forms of punishment—such as hanging—that were considered to be comparatively less painful. The painlessness argument persists up to the present day in the United States, where the lethal injection is considered as an intrinsically painless method for executing offenders. The lethal injection, as developed since 1977 in Oklahoma, consists in the administration of three drugs, the first of which, sodium thiopental, induces anesthesia in the offender. The goal is to render the offender unconscious, so that she does not feel the paralysis induced by the second drug (pancuronium bromide) or the intense pain caused by the third one (potassium chloride), whose function is to induce heart failure.

The historical argument is obviously insufficient if we want to show that capital punishment does not have to produce physical suffering. However, one can emphasize that both the supporters and the adversaries of the death penalty tend to agree that what renders this form of punishment problematic resides in its afflictive effects. Those who criticize capital punishment argue

punishment, see *Death Sentences and Executions 2013* (Amnesty International, 2014).

[37] Pieter Spierenburg, *Violence and Punishment: Civilizing the Body Through Time* (Cambridge: Polity, 2013), 114.

that all suffering that goes beyond the mere fact of execution is gratuitous and, as such, renders the death penalty morally problematic. Those who, on the contrary, defend capital punishment, claim that the experience of suffering is not problematic as long as it is not inherently contained in the fact of execution itself. Thus, representatives of both positions tend to dissociate between capital punishment and its potential afflictive dimension. This means that, at least at the level of jurisprudential argument, suffering is not taken to be a necessary or intrinsic feature of the death penalty, but a contingent and external aspect that is more or less stably associated with the way in which the death penalty happens to be administered.[38] To put it differently, the experience of physical suffering does not characterize capital punishment in the same way in which locking up the offender characterizes imprisonment.

The idea of a synthetic, non-essential relation between capital punishment and physical suffering is not new. As recalled by the *amicus curiae* offered by the *American Association of Jewish Lawyers and Jurists* in *Baze v. Rees* (2008),

> 2000 years ago the rabbis of the Talmud agreed that notwithstanding the apparent literal meaning of the text, *execution must be carried out as painlessly as possible*. The relevant passages from the Talmud demonstrate that the rabbis sought—with the scientific knowledge and the means available to them in their time—to formulate the quickest, least painful, and least disfiguring methods of execution that the technology of the day would allow within the framework of Biblical texts.[39]

[38] For a closer analysis of this implicit consensus, see Amanda Pustilnik, "Pain as Fact and Heuristic: How Pain Neuroimaging Illuminates Moral Dimensions of Law," *Cornell Law Review* 97, no. 4 (2012): 801-848.

[39] Quoted in Robert Blecker, "Killing Them Softly: Meditations on a Painful Punishment of Death," *Fordham Urban Law Journal* 30, no. 4: 974.

The thought here is not that all cases of capital punishment actually were or are painless, but that one should try to opt for those penal methods that minimize pain up to the point of eliminating it entirely. The very idea of trying to reduce or avoid pain as much as possible relies on the background assumption that there *can* be a painless form of capital punishment. If capital punishment without pain were not attainable, then the project of making it as painless as possible would be futile.[40] We have no good reason to think that the supporters of painless capital punishment are committed to penal futility.

In order to illustrate the possibility of painless capital punishment, we can resort to F.H. Bradley's case for construing legal punishment without resorting to the notion of physical suffering. Arguing against those who claim 'punishment consists in the infliction of pain for pain's sake,' Bradley takes the time to emphasize that

> Pain, of course, usually goes with the negative side of punishment, just as some pleasure, I presume, attends usually the positive side. Pain is, in brief, *an accident of retribution*, but certainly I never made it more, and I am not aware that I made it even an inseparable accident. If a criminal defying the

[40] One could think of other practices—say, medicine—where reducing pain as much as possible remains meaningful even if we cannot remove pain altogether. But there is at least one important difference between punishment and medicine in this respect: the pain caused by a medical intervention is meant to avoid a greater pain whose existence is not causally dependent on medical practice, whereas the potential pain of killing someone would be, in this case, a direct causal effect of the penal practice. Knowing that, absent my penal intervention, there is no significant risk that the offender will die in physical pain, I would be logically inconsistent and morally disingenuous in claiming that the rationale of my intervention is to remove the pain of the offender's dying altogether. This is because it is precisely my intervention that causes the existence of physical pain in the first place.

> law is shot through the brain, are we, if there is no pain, to hold that there is no retribution?[41]

More generally, one can argue that, if capital punishment is possible in the absence of physical suffering and capital punishment is an important instance of punishment, suffering is not a necessary feature of legal punishment. This argument can be pushed even further. Upon closer examination, we come to see that the death penalty cannot entail the suffering of the offender: a dead offender cannot be the subject of any experience, afflictive or otherwise. This is because, if capital punishment consists in the act of execution itself, then once the execution has been performed, the offender is no longer there to experience its effects. Conversely, capital punishment has not been imposed as long as the offender is still alive.[42] This singles out capital punishment as a form of punishment whose effects cannot be actually experienced by the punished.

The example of capital punishment undermines the distinction between core and peripheric instances of punishment. In particular, it shows that this distinction is unable to back up the afflictive definition. The supporter of the afflictive view could try to counter my answer in two ways. On the one hand, she could remind me that physical suffering does not exhaust the whole range of types of suffering. Indeed, as indicated in my analysis of the hard ambiguity problem, not all suffering is physical. On the

[41] F.H. Bradley, "Some Remarks on Punishment," *International Journal of Ethics* 4, no. 3 (1894): 284.

[42] The idea that the (potentially) afflictive effects of capital punishment cannot be experienced by the executed offender is also defended in Joseph Zelmanowits, "Is There Such a Thing as Capital Punishment?," *British Journal of Criminology* 2, no. 1: 78-81. However, Zelmanowits' argument is that, since capital punishment is, by definition, suffering-free, we should not consider it as a form punishment. On the contrary, a non-afflictive definition of punishment shows why we should count capital punishment as punishment.

other hand, the defender of the afflictive view could emphasize that capital punishment produces a significant amount of non-physical suffering, especially in terms of the psychological suffering undergone as the offender awaits execution.

Both of these arguments are definitionally inconsequential. The second argument is inappropriate because it conflates the suffering resulting from the actual imposition of punishment with the suffering the offender might undergo in anticipation of its enforcement. The latter form of suffering is not, properly speaking, a characteristic of punishment itself, just as the suffering that might be felt during a criminal trial is not a feature of the penal sanction potentially decided at the sentencing phase. The kind of suffering that goes with the trial derives from the possibility of punishment, not from punishment itself. Similarly, the constant suffering lived on death row is the suffering of the waiting, not that of undergoing the death penalty itself. This supplementary distinction between the possibility—or the probability—of punishment and its actuality also shows why the non-physical forms of suffering that can be experienced prior to the real imposition of capital punishment are irrelevant when it comes to defining punishment itself. The definition of legal punishment should be the definition of the act of punishment *tout court*, not that of its possible or probable—and thus contingent—effects.[43]

[43] Arguing that penal death represents an afflictive experience for the executed offender would require some plausible proof about the persisting existence of the person or entity undergoing this experience. Since, as the literature on posthumous interests shows, this proof remains at least controversial, saving the afflictive definition of punishment by positing the existence of a posthumously experiential person would come at an excessive metaphysical price.

Since the fine problem cannot be dismissed on the basis of a core-periphery distinction between different forms of punishment, the afflictive definition proves to be extensionally inaccurate. The afflictive definition is also inaccurate, insofar as it cannot make sense of those rarer instances of punishment where the offender does not or is unable to experience penal suffering. This is what might be called the masochist-callous offender problem.[44] The masochist offender is a person who takes pleasure in being subject to penal sanctions, whereas the callous one shows total indifference to punishment. Both cases raise a difficulty for the supporter of the afflictive definition. If suffering is not actually experienced, it is not clear whether, according to a suffering-centred view of punishment, the masochist and the callous offenders can be punished. Advocates of the afflictive view of punishment usually counter the masochist-callous objection by arguing that it is not the actual experience of suffering that matters, but the punisher's intention of imposing it.[45]

Nevertheless, separating the experience of the punished from the intention of the punisher is not as straightforward as defenders of the afflictive definition assume. It is difficult, if not impossible, to properly understand punishment if we choose to ignore what being punished means. In order to better see why this is so, imagine that the punisher knows that she is confronted with a case of penal masochism or callousness. If the punisher's intention is to inflict suffering, then she has a good reason to

[44] For an initial formulation of the masochist problem, see Tziporah Kasachkoff, "The Criteria of Punishment: Some Neglected Considerations," *Canadian Journal of Philosophy* 2, no. 3: 364-365.

[45] See, for example, John Kleinig, *Punishment and Desert* (The Hague: Martinus Nijhoff, 1973), 24.

adjust the penal sanction in a way that renders the masochist and the callous offenders negatively sensitive to it.

With respect to the masochist, the imposition of suffering can be realized by suspending the penal sanction, thus depriving the offender of her source of penal pleasure. However, this means that the punisher does not punish anymore in the legal sense of the word. The suffering of the offender is not an aspect of her being punished, but an effect of the punisher's intentional choice to refrain from punishing. In the case of the callous offender, the punisher could adjust the administration of the penal sanction up to the point where the offender is negatively affected by it. It is not clear how feasible such adjustments are. An offender who is really unresponsive to suffering-inducing actions will very likely require creative forms of punishment that will seldom be available in the institutionally constrained context of the criminal justice system. Such creative penal sanctions will conflict with the legality requirement that characterizes legal punishment. In any case, even if the callous offender could be made to suffer, this still does not solve the problem posed by the masochist.

Alternatively, the punisher can choose not to adjust the penal sanction at all. Doing nothing, however, cannot be adequately characterized as the intentional imposition of suffering, especially since the punisher knows that neither the masochist nor the callous offender will suffer. Compare this with a situation where a teacher is trying to instill knowledge on children following a set of methods whose pedagogic potential she positively knows is null or negative. If the teacher did nothing to change or replace those methods, could we still continue to describe her activity as the intentional imparting of knowledge on children? The two situations are similar in at least one respect: just as the teacher would be merely pretending to teach, a punisher who knows that

her actions are doing nothing to produce the suffering of offenders would be pretending to punish.

All this shows that the distinction between the experience of suffering by the offender and the intentional imposition of suffering by the punisher is unwarranted and does not adequately answer the masochist-callous offender problem. To sum up on this point, it would be more appropriate to say that a punisher who is dealing with a masochist or callous offender can hope or wish or wait for the respective offender to suffer. It is nonetheless inaccurate to affirm that the punisher *intends* to impose suffering on the offender, when she knows that the offender is, as a matter of fact, enjoying or reacting indifferently to her penal sanction.[46]

The Problem of Non-Operationalizability

The detailed analysis of the ambiguity and the inaccuracy problems should help us understand more quickly why the afflictive definition is flawed in a third way, namely, insofar as it is non-operationalizable. Defining punishment in terms of suffering is not useful if we want to use this definition as a basis for a subsequent operational definition of punishment. An operational definition is supposed to offer a validating test for performing those operations that are sufficient to construct the *definiendum* or assess whether we are in its presence. For example, the definition of a circle as the locus of points equidistant from a different, fixed point assists us in identifying the kind of operation—say, drawing a circle using a pair of compasses—that is suitable in order to construct it. An operational definition specifies an

[46] For an analysis of the distinctions between intending, wishing, hoping, and so on, see G.E.M. Anscombe, *Intention* (Oxford: Blackwell, 2nd ed., 1963).

observable condition or process for detecting the actual instances of a *definiendum*.

The reason why the afflictive definition cannot provide us with a suitable operational test for punishment is a direct consequence of the ambiguity of suffering. As already indicated, the four dimensions of ambiguity in suffering make it particularly hard, if not altogether impossible, to identify a method for guaranteeing whether we are punishing in the afflictive sense of the term or not. The discussion of the inaccuracy problem in its intensional dimension is illustrative of these operational limits. Insofar as the hard ambiguity problem is unsolvable, the afflictive definition will impede us in formulating *any* satisfying operational definition of punishment. On the one hand, this is because we can never be sure whether we are intentionally making offenders suffer in the relevant sense. On the other hand, there are cases—such as the one raised by the penally masochist or callous person—where it is highly uncertain whether we can punish at all. Therefore, the afflictive definition excludes punishment from the class of operationalizable *definiendi*. This should give us reasons to worry, especially if we think that punishment is required as an appropriate practical response to offences.

III

Toward a Non-Afflictive Definition of Legal Punishment

The ambiguity, inaccuracy and non-operationalizability of the afflictive definition are sufficient reasons for abandoning it. If you are convinced by the need to do so, then the goal of this article has been largely attained. Still, I need to say something about the way in which legal punishment could be defined without resorting to the idea of the intentional imposition of suffering on offenders. Additionally, it will be useful to highlight some of the

implications that flow from eliminating the afflictive definition. I will try to briefly address both of these issues in this third, and final section.

Sanctions, Not Suffering

One way of amending the afflictive definition would be to remove suffering as part of the *definiens* and keep the other four components enumerated at the beginning of the second section. This negative strategy is unsatisfactory, since it makes the definition incomplete. Saying that legal punishment is 'done on putative offenders for their offences by a person different from them and holding a special authority to do so' fails to indicate *what* is the kind of action punishment refers to. Something should be added to this statement before we can consider it as a candidate for definition. I suggest that the notion of *sanctions* is an adequate substitute for *suffering*. The idea of sanctions is normatively thinner than that of suffering, and it generally refers to the practical consequences legally attached to a particular action, activity or conduct. More specifically, sanctions are coercive measures whose authorization and enforcement are elicited by one's failure to comply with a specific law, rule or order.[47]

Modifying the first component of the *definiens* so that it becomes 'the intentional infliction of sanctions' escapes the three flaws of the afflictive definition. It avoids ambiguity, since the idea of sanctions can be positivistically construed as whatever counts as a sanction—or as its functional equivalent—within a particular jurisdiction. This also solves the accuracy and the operationalizability problems. If all forms of legal punishment are

[47] This is the sense in which the *Black's Law Dictionary* defines *sanction*. See *Black's Law Dictionary* (West, 9th ed., 2009), 1458.

sanctions and if sanctions are authoritatively spelled out in terms of the rules of a particular jurisdiction, it will be impossible to exclude paradigmatic forms of punishment from its definition. It will also be relatively simple to devise a procedure—say, reading the legally authoritative texts or consulting the relevant jurisprudence—to test whether we are dealing with a case of punishment or not. More generally, the terminology of sanctions is sufficiently capacious to accommodate any new or alternative forms of punishment—like criminal restitution, fines or community work—whose rationale does not seem to reside in their potential to make offenders suffer.

Sanctions should be understood enumeratively. This implies that their content is going to have the clarity of those authoritative rules—be they oral or written—that specify them. Furthermore, the idea of sanctions does not inherently rest on the offender's negative experience of punishment. The sanctioning character of an action is given by its bindingness or coercive character, not by its afflictive effects. The binding and coercive character of an action is neither necessary nor sufficient for something to be considered as afflictive. Taxes, for example, are both binding and coercive, and yet we do not try to define them in terms of the suffering that they necessarily generate.[48] On the

[48] At first blush, replacing *suffering* with *sanctions* collapses the distinction between civil, fiscal and criminal sanctions. This impression is mistaken. The distinction between these three kinds of sanctions persists in a legally positive way. This means that their difference resides in the fact of their formalization by different bodies of law. From a positivist standpoint, there are at least two important features that single out criminal sanctions. Unlike fiscal measures, the enforcement of criminal sanctions is practically incompatible with the continuation of the activity that is being sanctioned. Unlike civil sanctions, criminal ones are subject to stronger procedural safeguards. An account of punishment that is normatively thicker than the positivist one is not needed when it comes to its definition. My general argument in this article could be

other hand, a physical or psychological pathology can be afflictive, but it would be awkward to try to describe it as binding or coercive in the normative sense of the word. Therefore, 'bindingness' and 'suffering' are neither co-extensive nor co-intensive terms.

There are at least two objections that could be opposed to this definitional amendment. The first objection is that replacing 'suffering' with 'sanctions' makes the definition circular. The idea here is that 'sanction' is a loose synonym of 'punishment,' and that defining punishment in terms of sanctioning will thus make us run in a vicious circle. This objection is doubly misdirected. First, the meaning of sanctions is not strictly synonymous to that of punishment. The mere existence of civil sanctions testifies that not all sanctions are, by definition, penal ones. Second, the amended definition does not narrow down the *definiens* of punishment to its first component. Amending the afflictive definition does not modify or dismiss its other four components. These other components should be sufficient to indicate that the definition of punishment is not entirely circular. In particular, the requirement that punishment has to be imposed on a putative offender indicates that there is no extensional or intensional identity between punishment and sanctions. Sanctions *in general* are not necessarily inflicted on penally specified offenders.

The second objection is that eliminating suffering from the definition of legal punishment will render us insensitive to the afflictive effects of penal practice. Call this the objection from concealment. The objection claims that taking suffering out of the definition of punishment will result in obscuring, hiding, and covering up the fact that penal sanctions are often the sources of the offenders' negative experience, like physical and psychological

understood as asserting that we should be positivists about the definition of punishment, even if we do not have to be positivists about its justification.

pain, social or moral harm, as well as other various forms of unpleasantness. To put it in Scott Veitch's terms, a non-afflictive definition of punishment would simply illustrate, once again, the persistence of the 'amnesiac capabilities of legal thought and practice in the legitimation of human suffering.'[49] This objection is also doubly misguided. First, it conflates the definitional task—which is to bring clarity and accuracy into our understanding of the *definiendum*—with a normative one, which is to justify, criticize or practically modify the actions that fall under the *definiendum*. Second, the objection from concealment ignores that including suffering in the *definiens* of legal punishment cannot by itself serve as a safeguard against its potential afflictive effects. The objection can be thus turned on its head. Expecting punishment to always involve suffering might also normalize it, thus making us insensitive to it. On the contrary, a non-afflictive construal of punishment could be interpreted as a way of creating the conceptual space that is needed to regard suffering as a contingent—and thereby avoidable—aspect of penal practice.[50]

Two Implications of the Non-Afflictive Definition

Most of the article focused on the definitional reasons for abandoning the afflictive understanding of punishment. I would like to briefly end by highlighting two important normative implications of the non-afflictive definition. The first implication is that opting for a non-afflictive definition levels the playing field between different—and potentially mutually exclusive—

[49] Scott Veitch, *Law and Irresponsibility. On the Legitimation of Human Suffering* (Abingdon: Routledge-Cavendish, 2007), 4. See also Linda Ross Meyer, "Suffering the Loss of Suffering: How Law Shapes and Occludes Pain," IN *Knowing the Suffering of Others. Legal Perspectives on Pain and its Meanings*, ed. Austin Sarat (Tuscaloosa: The University of Alabama Press, 2014), 14-61.

[50] For a more detailed response to this argument, see Coverdale, *op. cit.*

justifications of punishment. In particular, it allows supporters of retributivism to defend their position in a way that does not make their justification otiose. If, as Quinton used to argue, retribution is a logical doctrine, inasmuch as 'suffering for suffering's sake' represents an 'elucidation of the word ['punishment'],' then any conception that takes the offender's suffering to be a ground for punishing will merely restate what it means to punish.[51] It will not give us a distinct normative *reason* to punish. This does not imply that the non-afflictive definition is devised to save retributivism from normative irrelevance. Taking suffering out of punishment makes equal room for other conceptions as well, like restorative justice or more consequentialist views of punishment. If restorative justice is radically opposed to the offender's suffering and the utilitarian-minded consequentialist is committed to reducing (useless) penal suffering as much as possible, then only a non-afflictive definition allows us to make sense of these theories as theories of punishment proper. To put it more generally, the non-afflictive view of punishment remains neutral in relation to different normative accounts of legal punishment.

The second implication is closely connected to the first one. Adopting a non-afflictive understanding of legal punishment shows that penal abolitionism rests on a definitional mistake. There are different forms of penal abolitionism, but the most vigorous ones—like the one held by Louk Hulsman or, more recently, Daniel Boonin and Nathan Hanna—claim that punishment should be abolished precisely because it consists in the intentional infliction of suffering on offenders.[52] Abolitionists take suffering to be a necessary feature of any penal practice. If

51 A.M. Quinton, "On Punishment," *Analysis* 14, no. 6: 134.

52 Louk Hulsman, "The Abolitionist Case: Alternative Crime Policies," *Israel Law Review* 25, no. 2-4: 681-709; Boonin, *The Problem of Punishment*; Hanna, "Liberalism".

this is not the case, then abolitionists lose their main rationale for rejecting punishment.

There is a sense in which the strength of the abolitionist project is predicated on the indistinct use that its supporters make of the terms *penal* and *punitive*. The meaning of these terms is not the same. The terms *penal* and *punishment* denote a specific type of legal sanction, whereas the terms *punitive* and *punitiveness* connote—and tend to criticize—the normatively unacceptable content or enforcement modality of a legal sanction. Oftentimes, abolitionists talk about punishment as if it were a synonym for punitiveness. More specifically, they tend to equate punishment with imprisonment, argue that imprisonment is essentially punitive, and conclude that punishment is essentially punitive and, as such, radically unjustified.[53] Since the first premise is false, the whole abolitionist argument is unsound. Though prisons tend to function punitively, they do not exhaust the entire range of penal sanctions. As already indicated, other sanctions, such as fines, criminal restitutions or community work are penal, but this does not make them inherently punitive.[54]

A non-afflictive position would allow contemporary supporters of abolitionism to concentrate on the issues that really bother them, which are punitiveness and penal excess.[55] Appropriately understood, the problem that lies at the basis of abolitionist attitudes is the suffering that comes out of certain forms of punishment as practiced today, and not the fact of

[53] For an example of such a slippery use, see Howard Zehr, *The little book of restorative justice* (Intercourse, PA: Good Books, 2002).

[54] For a detailed critique of the ambiguous use of the term *punitive* and *punitiveness*, see Roger Matthews, "The myth of punitiveness," *Theoretical Criminology* 9, no. 2: 175-201.

[55] For an analysis of penal excess, see David Garland, "Penal Excess and Surplus Meaning: Public Torture Lynchings in Twentieth-Century America," *Law & Society Review* 39, no. 4: 793-834.

punishing itself. Thus, for example, an abolitionist should not have a problem with penal sanctions that take a compensatory or restitution-based form and that are not deliberately directed at the offenders' suffering. Given that such sanctions actually exist *qua* penal sanctions, the abolitionist would have at least a *prima facie* reason to accept them on the basis of their non-afflictive content.[56] Doing so, however, would force the abolitionist to hold an inconsistent view, since she would, on the one hand, reject punishment overall and, on the other hand, be compelled to endorse certain existing non-afflictive forms of punishment. This internal inconsistency makes abolitionism ultimately untenable.

To summarize, we have three good reasons to give up on the afflictive definition of legal punishment, namely, its ambiguity, inaccuracy and non-operationalizability. We also have reasons to be optimistic about the possibility of an alternative, non-afflictive definition of punishment. This latter definition avoids the flaws of the former one. It also produces two interesting normative consequences. The first consequence is that we have a more neutral basis upon which different justifications of punishment can engage in principled deliberation. The second one is that a non-afflictive understanding of punishment allows us to see why penal abolitionism is a definitionally misdirected project. More generally, and though I take the sanction-based view of punishment to be a sound one, I believe that a discussion about

[56] For example, in the U.S., restitution is introduced as a penal sanction by the *Victim and Witness Protection Act* in 1982. For an in-depth analysis of penal restitution, see Charles F. Abel, Frank H. Marsh, *Punishment and Restitution. A Restitutionary Approach to Crime and the Criminal* (Westport, CT, London, England: Greenwood Press, 1984). For a positive analysis of criminal restitution, see Courtney E. Lollar, "What is Criminal Restitution," *Iowa Law Review* 100: 93-154.

the appropriate definition of legal punishment deserves to be addressed anew.[57]

Université de Montréal

Centre de recherche en éthique (CRÉ)

[57] I would like to thank the two anonymous reviewers for their insightful and highly instructive comments. I would also like to thank the organizers of the 2014 European Conference of Analytic Philosophy, where an earlier draft of this article was presented. My special acknowledgments go to Daniel Putnam and Elise Rouméas for their careful reading and critical comments of the initial version of this article.

If you need to cite this article, please use the following format:

Poama, Andrei, "Punishment Without Pain Outline for a Non-Afflictive Definition of Legal Punishment," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), 97-134, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

THE PHILOSOPHY OF PUNISHMENT

CRIMINAL LAW AND THE INTERNAL LOGIC OF PUNISHMENT

BY

KATRINE KRAUSE-JENSEN & RAFFAELE RODOGNO

 Philosophy and Public Issues (New Series), Vol. 5, No. 1 (2015): 135-171
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

Criminal Law and the Internal Logic of Punishment

Katrine Krause-Jensen & Raffaele Rodogno

We argue that punishment has an essentially retributive core that carries its own retributive type of logic or reasons. In particular, we show that punishment is something that we understand as in principle always being assessable in terms of deservingness and that this is ultimately to be understood in terms of moral culpability and nothing else. These features make up what we call the internal logic of punishment. The practice of punishment, however, can also be assessed with a logic that is external to it. What this consists in is first and foremost determined by the aims and constraints of the punishing agent. For the modern liberal state these are typically understood in terms of deterrence, incapacitation, rehabilitation, and, arguably, the expression of condemnation. The idea that punishment has its own internal logic has a number of consequences with regard to criminalization, to the extent, that is, that the latter involves punishment. For one, purely instrumentalist justificatory accounts of punishment will not work as they fail properly to consider the retributive core of punishment. Next, we consider what follows from the fact that by inflicting punishment, the state takes it upon itself to mix these two logics, the internal and the external, together. In particular, we bring forward some tensions that arise when the state mixes

Philosophy and Public Issues (New Series), Vol. 5, No. 1 (2015): 135-171
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

the internal logic of punishment with certain modern, liberal aims and constraints that are external to punishment.

I

The Question and its Method

The criminal law, with its emphasis on punishment, is generally assumed to perform a number of functions. Deterrence and incapacitation are the most obvious ones. One way to envisage the criminal law is as an attempt to regulate the behaviour of those who inhabit its territory through deterrence and incapacitation. State punishment is the particular form of deterrence and incapacitation that behaviour regulation takes when handled by the criminal law. State punishment, however, is not the only mode of deterrence and incapacitation. The state may for example impose non-punitive sanctions in order to deter certain forms of behaviour, or restrain a person to an isolated space in order to disable her from spreading a dangerous disease. Given this fact, one legitimate question arises: is the criminal law simply another mode for the regulation of behaviour in the hands of the state or does the criminal law, with its emphasis on punishment, have a distinctive character or distinctive aims? What answer one gives to this question is important insofar as it will favour or disfavour answers to another urgent question: what kinds of conduct should be subject to the distinctive mode of control that is the criminal law?[1]

[1] R.A. Duff, L. Farmer, S.E. Marshall, M. Renzo, and V. Tadros "Introduction," in R.A. Duff, L. Farmer, S. E. Marshall, M. Renzo, and V. Tadros (eds.), *The Boundaries of the Criminal Law* (Oxford: Oxford University Press 2010), p.6.

Most broadly, the relevant theoretical landscape can be divided between, on the one hand, *instrumentalists* who see the criminal law as *just* another tool for the regulation of behaviour in the hands of the state with no distinctive character or aims of its own; and, on the other, *non-instrumentalists* who take it that the criminal law, with its emphasis on punishment, does have distinctive character and/or aims. Of course, a more fine-grained curving of the conceptual space brings forward important distinctions that somewhat soften the contrast between instrumentalists and non-instrumentalists. This is for example achieved by hybrid views that admit both instrumentalist and non-instrumentalist elements such as Duff[2] and, to a lesser extent, by non-consequentialist, instrumentalist views such as Tadros.[3] In this paper, however, we concentrate on what separates instrumentalists from non-instrumentalists rather than on what unites them, and what does separate them in the end is their respective stance on the question of the distinctiveness of the criminal law.

Note that though the question sounds descriptive, many philosophers are inclined to read it in normative terms. In other words, even if the question asks what, if anything, *is* distinctive of the criminal law, philosophers typically understand it as asking what *ought* to be considered as distinctive or, better, what would be distinctive of the criminal law on an *ideal* account of the latter. While we are also ultimately concerned with this philosophical understanding of the question, in this paper our main concern is to elucidate one of the fundamental concepts of the criminal law, i.e., punishment, and to do so *not* by offering another ideal

2 R.A. Duff, *Punishment, Communication, and Community* (New York: Oxford University Press 2001).

3 V. Tadros, "Criminalisation and Regulation," In R.A. Duff, L. Farmer, S. E. Marshall, and M. Renzo (eds.), *The Boundaries of the Criminal Law* (Oxford: Oxford University Press 2010).

account of the latter but by an interpretation of punitive practices that emphasizes their psychological underpinnings. As we are to explain, such analyses do have some normative import and are hence relevant to the philosophical question at hand.

The analysis of punishment that we present below is in family with historical and interpretative analyses of the criminal law (Lacey 2009). Unlike standard conceptual analyses, these analyses are not aimed at straightening out inconsistencies through systematization. They start by identifying the (disparate) elements of a certain concept or practice (e.g. the criminal law) in order to offer an understanding of its current form by illustrating the circumstances in which it came about, the purposes that it served in those circumstances, and how it evolved, survived, and adjusted to new contexts. The analysis we intend to offer can be seen as complementing this type of interpretative, historical analysis by focussing on some of the psychological features that cannot be ignored in order to achieve a correct understanding of our concept of punishment and the practices that it animates, such as the criminal law. At the most basic level, the idea behind this type of analysis is that practices such as punishment (in its various forms) are not merely the product of different cultural contexts but also the expression of specific psychologies. Creatures whose psychological profiles significantly differ from ours will likely develop practices and concepts different from ours. Hence, for example, in a world made of creatures who unfailingly obey the law, while it may still be necessary to legislate in order to co-ordinate behaviour, coercive practices such as criminal punishment may be superfluous and would in all likelihood fail to arise.[4]

[4] J. Finnis, *Natural Law and Natural Rights*, 2nd ed. (Oxford: Oxford University Press 2011), pp. 269-270.

Clearly the approach described here is not as such a normative approach. Yet, if correct, analyses that embody these approaches do impose constraints on ideal accounts of the criminal law. This is not because our approach hinges on the controversial assumption that we can derive an 'ought' from an 'is'. That this is not the case can be shown by looking at the type of constraints that we have in mind and the way in which they interact with normative or ideal theories. The constraints we have in mind are of two kinds.

Firstly, consider this. Ideal or normative accounts of the criminal law are not at freedom to provide any account of punishment whatsoever but must provide accounts in which punishment can still be recognized as such. By contributing to an analysis of punishment, we set the frame, and hence the limits, within which any normative or ideal account of the criminal law can legitimately operate, if it is to be recognizable as such.

Secondly, the type of analysis provided below is substantive. In particular, we will show that punishment is retributive at its core. This is to say that, if normative or ideal accounts of the criminal law are to justify punishment at all, they must justify a practice that has some distinctive retributive elements. This may set significant constraints on the normative theorist, for, as we will see, justifying retributive punishment may be especially arduous in the context of liberal politics. Yet this is not to say that the best normative or ideal account of the criminal law must be at least partly retributive, as one may of course be abolitionist and/or defend as non-punitive interventions as the only justified form of legal regulation of behaviour.

Given the centrality of punishment to the (disputed) distinctiveness of the criminal law, we will dedicate the bulk of the paper to the analysis of punishment and then examine its implication for the criminal law. More in particular, we begin by

focussing on generic punishment (Section 2) on the assumption that legal punishment is indeed a specific form of generic punishment.[5] To anticipate a little, we argue that punitive action of any kind is something that originates in our emotional life and in particular in our sense of justice and the emotions that are in family with anger (Section 3 and 4). Punishment, as a concept and a practice, has a distinctive emotive logic that involves distinctive retributive and condemnatory features. Any punitive practice worth of its name cannot escape this fact (Section 5). With this understanding of punishment in hand, we return to the dispute between instrumentalists and non-instrumentalists (Section 6 and 7).

II

Punishment

It is standard fare for legal theorists to start one's account of punishment with the claim that it consists in the infliction of some burden, deprivation, harm, or hard treatment or more generally, of a *disvalue*, or something that a given community recognizes as such. While this is indeed an essential feature of punishment more in general, it would be misguided to think that it is all there is to it. In particular, punishment is essentially historical. To illustrate this consider an act that consists in the intentional infliction of pain on someone. Whether this act amounts to an instance of punishment or an instance of assault will essentially depend on what comes before it. We claim that for it to count as an instance of punishment, it must be understood

[5] This should be an uncontroversial assumption, as it only claims that nothing is to count as an instance of punishment, legal or otherwise, unless it displays those features that are agreed to be necessary for anything to count generically as punishment.

as a *reaction* or a *response*. More in particular, punishment is the infliction of a disvalue as a reaction or response to a *perceived injustice, wrongdoing*, or *violation*. Any infliction of harm that cannot be understood as a reaction of this kind is likely to be understood as an assault, a wrong, or a violation of some kind rather than as a punishment.

Envisaging punishment as a reaction to perceived wrongdoing has one important consequence. In reacting to an (alleged) misdeed by inflicting a disvalue on the alleged wrongdoer, it is quite clear that we are sending the message back to him that his action was unwelcome. This is not an accidental feature of punishment: we want (perceived) wrongdoers to undergo something negative as a response to the wrong they are perceived as having committed. This does indeed come very close to the idea that punishment is in its nature (rather than in its justification) condemnatory.

Punishment then is essentially a reaction to perceived wrongdoing. But it is also more than simply that. It is an undeniable feature of our practice that it comes in unmistaken normative language, the language of justice and deservingness. Whether punishment is just is in general evaluated in terms of desert.[6] Punishment is something the innocent does not deserve, something that the wrongdoer deserves, and, in fact something that he deserves to greater or minor extent depending on the gravity of his deed. In other words, punishment is always

[6] Note that the deservingness relation between the (punishable) act and its (punitive) response does not on its own determine whether it is right or just all things considered to inflict punishment (and what kind and amount of punishment) nor whether there is most reason or it is most rational to inflict punishment (and what kind and amount of punishment). In short, the question of the deservingness of punishment is separate from the question of its infliction (and as we will argue later regulated by distinct 'logics').

considered to be either deserved or undeserved and when deserved there are concerns about the amount, kind, or intensity of punishment that the wrongdoer deserves.

This normativity, however, should not be understood as a sheer linguistic feature of punitive practices. The nature of our discourse should rather be taken to express the phenomenology of our punitive responses. When we punish, or are undergoing the impulse of doing so, our perception of a wrong, and unjustified harm or slight, is accompanied by a feeling of injustice. What is more, we tend to feel entitled to, justified in, or righteous about our punitive attitude. If we perceive that the offender has gotten what he deserved, our feeling of injustice will subside: justice has been done. But if we perceive that the offender has "gotten away with it", the sense of injustice—feelings that the state of affairs ought to be rectified, that the offender deserves to pay for his misdeeds—will linger on for some time.

The immediate reaction that someone deserves punishment is generally modulated by two factors: wrongdoing (which could also amount to an omission) and moral responsibility or culpability. If you committed a wrong act but were not at all culpable (you killed someone while unwillingly hypnotized), no one would say that you deserved to be punished for it. A combination of these two factors, e.g., the severity of the wrong committed and the degree of culpability determines our thoughts about the amount, kind, and intensity of punishment that is deserved.

It becomes clear from these features that our reactive punitive attitudes are *retributive* in nature, for retribution is nothing other than the idea that one should get what one deserves, where this is uniquely determined by the culpability and severity of one's perceived wrong. Given our purposes, this is an arresting conclusion, as it implies that the criminal law has at its core a

practice whose retributive and condemnatory nature would render sufficiently distinct from other modes for the regulation of behaviour available to the state, whose nature is neither retributive nor condemnatory. Instrumentalism would at this point begin to look simply off track, as a view offering an account of something other than punishment.

This central conclusion, however, is in need of greater argumentative support. After all, one may object that the argument so far consists in a mere appeal to intuition and phenomenology. What we would have so far is an arbitrary collection of claims about punishment that include, among others, the idea that it is retributive. In section 3, we will therefore attempt to show that the characterization of punishment provided here is not a mere collection of disparate features but that of a unitary phenomenon. On our account, what gives unity to these features is the fact that punishment is, in some sense to be explained, based on our sense of justice and the emotions in family with anger. In section 4, we show how through this account we gain evidence to the effect that punishment is indeed retributive.

III

The Emotive Account of Punishment

Consider once again our characterization of punishment as a reaction to the perception of a wrongdoing or injustice accompanied by feelings of injustice, and followed by the intentional infliction of disvalue on the perceived wrongdoer. When punishment is considered, more holistically, in these terms, it mirrors important elements of our sense of justice and the emotive basis to which this is often associated. Providing support

for this claim will be our main concern in this and the next section.

Consider the way in which we take norm violations to differ in kind. Norm violations come, as it were, in different colours. More specifically, while we may cognize some norm violations as *injustices* (or *wrongdoings*), we may cognize other violations as *imprudent* or *impolite* actions, i.e., the violations of, respectively, prudential norms (e.g., getting drunk the night before an important exam), or of norms of etiquette (e.g. burping loudly at the dinner table). The capacity to cognize norm violations as injustices is part of what we shall call the *sense of justice*. As we saw, this type of cognitions are integral to understanding any action as an instance of punishment as opposed to a simple act of aggression. There is, however, more to the sense of justice. The idea is that, most often, perceived injustice does not leave us cold but is rather accompanied by distinctive feelings and action tendencies. These perceptions are in other words intimately connected to our capacity to react emotionally and, in particular, to react with anger and related emotions such as resentment, indignation, outrage or fury, and moral outrage.[7] Let us henceforth refer to these emotions as a group as *justice-related emotions*. The thesis we propose is roughly that our concept of

[7] With regard to the specific nature of this connection, R. Rodogno "Robots and the Limits of Morality," in M. Nørskov (ed.), *Social Robots: Boundaries, Potential, Challenges* (London: Ashgate 2016), pp. 39-55, argues in favor of a constitutive claim to the effect that it is precisely these emotions that enable us (developmentally and phylogenetically) to cognize certain norm violations as injustices or wrongs. Note that this thesis is compatible with cases in which a subject perceives injustices unaccompanied by the relevant emotive reactions. For the purposes of this paper, however, it will not be necessary to assume this constitutive connection; a less controversial thesis to the effect that the sense of justice and the anger-related emotions are connected in some way (causally, statistically, or constitutively) will do.

punishment and the practices that it animates are phylogenetically and developmentally dependent on the sense of justice and the capacity to experience these emotions.

We can begin articulating and defending this thesis, by considering some classical characterizations of anger. On Aristotle's much discussed account, for example, anger is an impulse, accompanied by pain, to a conspicuous revenge for a conspicuous slight directed without justification towards what concerns oneself or one's friends.[8] Somewhat similarly, Roberts argues that in anger we construe the situation in these terms:

> S has culpably offended in the important matter of X (action or omission) and is bad; I am in a moral position to condemn; S deserves (ought) to be hurt for X; may S be hurt for X.[9]

From these classic examples, there appears to be a similarity between the characterization of punishment, on the one hand, and that of anger, on the other. First, just as punishment is a reaction to perceived wrongdoing or injustice, anger involves cognitions to the effect that someone has culpably (and hence unjustifiably) offended or attacked you and yours, or violated an important norm.[10] Second, just as in punishment we take

[8] Aristotle, *Rhetoric,* in *Complete Works. Revised Oxford Translation*, vol. 2. edited by J. Barnes (Princeton, NJ: Princeton University Press 1984), 1378a31-1380a4.

[9] R. Roberts, *Emotions, An essay in Aid of Moral Psychology* (Cambridge: Cambridge University Press 2003), p. 204.

[10] It must be noted that the psychological literature is actually divided on one aspect whose presence is decisive to the argument to come. Unlike the account that we will offer, neo-associationist accounts such as those found in L. Berkowitz and E. Harmon-Jones—"Towards and Understanding of the Determinants of Anger," *Emotion* vol. 4, n. 2 (2004), pp. 107-130—dismiss the idea that *other-accountability* and *unfairness* would necessarily characterize the formal objects of anger or, as they would rather say, that they are necessary "determinants" of anger: when angry, it does not follow that we perceive or cognize someone as culpable of an unfair action or wrong. On this account,

ourselves to be righteous in wanting to inflict a disvalue on the offender, in anger we take ourselves to be in a moral position to condemn. Third, in both cases we take it that the offender *deserves* to receive some disvalue. Fourthly, while anger typically involves action tendencies to the effect that the wrongdoer be hurt or be

---

the perception that one's goal is being frustrated or averted may on its own give rise to anger. Given the centrality of culpable wrongdoing (or unjust intentional behaviour) to our idea of punishment, anger understood in this sense would not qualify as a good explanation for it. Psychologists, however, are divided on this issue as attested by the appraisal approach championed by C.A. Smith and L.D. Kirby, "Appraisal as a pervasive determinant of anger," *Emotion* vol. 4, n. 2 (2004), pp. 133-138). In a recent study involving 832 high-school subjects, Kuppens et al. (2007)—P. Kuppens, I. Van Mechelen, D.J.M. Smits, P. De Boeck, and E. Ceulemans, "Individual differences in patterns of appraisal and anger experience." *Cognition and Emotion* vol. 21 (2007), pp. 689–713—have shown that while goal frustration is always a necessary determinant of anger, a large number of subjects will not experience anger unless they perceive the situation as involving other-accountability and unfairness. In other words, due to individual differences with regard to a number of dispositional traits, while goal-obstacle is necessary and sufficient for many, it is necessary but insufficient for many others who also need to appraise the situation as involving others-accountability and unfairness. Importantly, the authors further found that if a situation is perceived as involving these three elements – goal-frustration, other-accountability, and unfairness— almost all participants reported experiencing anger. Some may find it hard to identify, for example, the anger you feel when cheated by someone and wanting justice to be done with the emotion that you feel when inattentively tripping on a table leg and wanting to kick the table. We would find it natural to understand the first as an instance of anger at the wrongdoer and the second as an instance of irritation or frustration. For our purposes, however, we need not decide which camp, neo-associationism or appraisal theory, is right about this. Instead, we will call the reader's attention to the fact that anger is here designated as being in family with emotions such as resentment, indignation, outrage or fury, and moral outrage, all emotions that non-controversially involve other-accountability or the attribution of intentionality. We will simply take the type of anger relevant to our emotive account of punishment to be similar in this respect to other emotions that are in family with it.

inflected some disvalue, punishment involves the actualization of these tendencies. Finally, though this is not Aristotle's claim, we could articulate Aristotle's remark on the painful nature of anger by saying that the negative feeling at issue here is that which we feel when we perceive that an offender has not paid his due, and "justice was not done".

Now we take it that these similarities between anger, on the one hand, and our understanding of punishment, on the other, are not an accidental matter. We rather take it to suggest the thesis already mentioned above according to which:

> *Thesis*. The sensitivity to injustice and the capacity to react emotionally thereupon is necessary to understand the concept of punishment and the practices that it animates.

The claim here is that in the absence of the relevant emotive basis there would be no logic or intuitiveness in the flow from evaluations of culpability to inclinations to inflict disvalue, the flow embodied in justice-related emotional processes presented above. How could the idea of responding to perceived harm with the infliction of harm (or disvalue) be intelligible to anyone deprived of the capacity of such emotional processes? Suppose, for a moment that we could establish beyond doubt that a punitive social practice that responded to harm with harm could be shown to have no deterrent or educational effect whatsoever. While some (but by no means all) of us may thereby take such a practice to lack justification, most of us will still understand the practice, find it intuitive or intelligible, or displaying a certain logic or point. This intelligibility, we claim, is due to our sense of justice and our capacity to experience justice-related emotions.

The thesis we propose is not to the effect that each and every occurrence of punishment requires a corresponding occurrence

of anger in those who impart the punishment. That thesis would be quickly rejected by the existence of institutionalized forms of punishment, such as punishment by the state, in which state officials will often inflict punishment without necessarily feeling justice-related emotions. Our thesis does not try to establish a one-to-one connection between occurrences of punishment, on the one hand, and occurrences of justice-related emotions, on the other. It is a thesis about the genesis and, from there, the nature of the concept of punishment and the social practices that it animates.

The thesis is rather to the effect that grasping the concept 'punishment' requires certain emotional capacities because this concept is itself a product of these capacities. If correct, it would follow from this view that those individuals (human or otherwise) who lack this capacity or whose capacity is damaged or impaired will fail to understand our punitive practices. Individuals are born in social contexts that include punitive practices whose content was actively and progressively shaped through the ages by our ancestors' sense of justice and justice-related emotions. These practices, with their specific content, are already in place whenever any individual is born. As they develop their social, affective and cognitive skills, individuals come to learn about and understand the ambient punitive practices. Those individuals (humans or otherwise) unequipped with the relevant capacities will struggle to make sense of them.

In the remaining part of this section, we present and articulate five auxiliary theses, which, if true, would lend inductive support to *Thesis*. Whether there is indeed any evidence in favour of these auxiliary theses will be discussed in the next section.

Consider first:

> *Thesis 1.* Individuals who lack the capacity to feel justice-related emotions (or whose capacity has been significantly hindered or damaged) exhibit significantly different patterns of behaviour as a response to injustice.

Thesis 1 is indeed very close to our original thesis. However, while the latter focussed on the connection between the capacity to feel justice-related emotions and the *intelligibility* of the concept of punishment, the former focuses on the connection between that capacity and actual patterns of punitive *behaviour* (or lack thereof). Behavioural patterns will be taken as evidence that the correct type of understanding is in place. In particular, evidence to the effect that those whose relevant capacities are absent or impaired do not punish as much, or as hard, or at all, is indirect evidence to the effect that they lack the proper understanding of punishment and, hence, of the practices that it animates. Unfortunately, however, this thesis has to be laid to rest here, as no empirical evidence either in favour or against it seem to have been gathered to date.

With the next four theses, we shift focus from the *capacity* to feel justice-related emotions in general, to the way in which sensitivity to such emotions affects instances of punishment, and the way in which occurrences of the former affect and co-occur with occurrences of the latter. Hence:

> *Thesis 2.* Occurrences of justice-related emotions partly determine who and what is to be considered as deserving of punishment.

Clear evidence that the occurrence of justice-related emotions in a subject has an effect on the subject's judgements about who and what is deserving punishment is here taken as indirect

evidence in favour of our main proposal. The same goes for the next claim:

> *Thesis 3*. Individuals' sensitivity to justice and justice-related emotions affects the content, strength and frequency of their punitive attitudes.

You may be sensitive to justice in the sense that episodes that would typically not elicit for example anger in others, do elicit anger in you because you conceive that episode as unjust. And similarly, you may be more sensitive to justice than others in the sense that you typically experience justice-related emotions more intensely than them. As we understand the thesis, one should expect those who tend to feel for example anger more intensely (with regard to certain kinds of violations, or perhaps with regard to violations more generally) to hold harsher punitive attitudes. This thesis can be understood both at the level of single individuals or of entire groups, be they defined by culture, gender, or both. Hence it may be that due to certain cultural contingencies certain violations are experienced as for example more angering by certain groups as opposed to others. We should thereby expect these violations to be the object of harsher punitive attitudes on behalf of members of such groups.

These theses explicate three important senses in which punishment is based in justice-related emotions. Even though none of these theses implies that we must be undergoing an occurrence of justice-related emotions in order to experience punitive attitudes, the way in which we envisage punishment to be based in justice-related emotions would certainly involve the following:

> *Thesis 4*. Occurrences of punitive attitudes tend to co-occur with justice-related emotions.

The idea here is that punitive tendencies are a part of occurrences of justice-related emotions. Hence, occurrences of the latter will carry punitive attitudes in their stride. The inverse is also true but only typically so given that, as explained above, we may well make judgements about punishment in the absence of an occurrence of anger or other anger-related emotions.

Finally, given our understanding of punishment as inherently retributive, and given *Thesis 4*, we should expect the following final claim to be true:

> *Thesis 5*. Justice-related emotions co-occur with punitive attitudes that are retributive in nature.

As we are about to see, work in social psychology does indeed provide some evidence in support of theses 2-5, and thereby supports our claim that punishment is an emotionally based retributive concept closely connected to our sense of justice. In the next section, we provide a quick overview of the literature relevant to establishing this claim.

IV

The Psychology of Punishment

Social psychologists have begun to explore both people's explicit beliefs about the justification of punishment and their motivation in punishing. A number of studies is taken to show that ordinary people, while overtly declaring to punish on consequentialist as well as retributive grounds, in practice would tend to judge the appropriateness of punishment on various cases pretty much in accordance with retributive intuitions about just

deserts.[11] There is in other words a disconnection between what people take to be the justification of punishment and the way in which they tend to punish in practice.[12] Importantly, however, it looks like when deciding what punishment to impart on someone, individuals are driven by retributivist as opposed to consequentialist considerations.

Some of these studies measured and emphasized the presence and co-variation of anger in connection with the severity of punitive attitudes. Carlsmith *et al.*, for example, found that moral outrage ratings were a strong predictor of judgements about punishment and mediated the influence of retributive considerations on those judgments: seriousness of wrongdoing

[11] See, for example, R.M. McFatter, "Sentencing strategies and justice: effects of punishment philosophy on sentencing decisions," *Journal of Personality and Social Psychology* 36 (1978), pp. 1490–1500; R.M. McFatter, "Purposes of punishment: effects of utilities of criminal sanctions on perceived appropriateness," *Journal of Applied Psychology* vol. 67 (1982), pp. 255–267; D. Kahneman, D. Schkade and C. R. Sunstein, "Shared outrage and erratic awards: the psychology of punitive damages," *Journal of Risk and Uncertainty* 16 (1998), pp. 49–86; J.M. Darley, K.M. Carlsmith, and P.H. Robinson, "Incapacitation and just deserts as motives for punishment," *Law and Human Behavior* 24 (2000), pp. 659-683; K.M. Carlsmith, "The roles of retribution and utility in determining punishment," *Journal of Experimental Social Psychology* 42 (2006), pp. 437–451; K.M. Carlsmith, "On justifying punishment: The discrepancy between words and actions," *Social Justice Research* 21 (2008), pp. 119–137, K.M. Carlsmith and J.M. Darley, "Psychological aspects of retributive justice," *Advances in Experimental Social Psychology* 40 (2008), pp. 193–236; K.M. Carlsmith, J.M. Darley and P.H. Robinson, "Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment," *Journal of Personality and Social Psychology* 83 (2002), pp. 284–299; J. Baron and I. Ritov, "The role of probability of detection in judgments of punishment," *Journal of Legal Analysis* 2 (2009), pp. 553-590.

[12] Nadelhoffer et al challenged the evidence to the effect that we are retributivist in practice as gathered by the studies mentioned above but do provide evidence to that effect through another experimental set up.

and absence of mitigating circumstances tended to co-vary with reported anger and judgements on severity of punishment.[13] This we shall take as evidence in favour of Theses 2, 3, 4 and 5.

Unlike the studies mentioned so far, the studies that we are about to review share the feature of manipulating anger directly. Psychologists working in this area are usually interested in documenting co-variation and causal relations. As a result, they tend to conceptualize anger and punitive judgements/attitudes as separately operationalizable occurrences whose relation needs to be documented. As we shall see, while leading to interesting and useful results, this approach has limited exploratory power.

A useful study is Lerner *et al.*, in which anger was induced in some subjects but not in a control group in order to compare the two groups' respective punitive reactions.[14] More in particular, the experimenters induced anger by exposing individuals to a video displaying bullying behaviour. They then asked the subjects to rate the degree to which perpetrators of hypothetical harms (unrelated to those shown in the video) should be punished. The punishment ratings for the subjects in the anger induction group were higher than those for subjects in a control group, indicating that incidental anger has a causal effect on (spills over) judgments about punishment, in line with (some version) of Thesis 4.[15]

13 K.M. Carlsmith, J.M. Darley, P.H. Robinson, "Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment."

14 J.S. Lerner, J.H. Goldberg, P.E. Tetlock, "Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of responsibility," *Personality and Social Psychology Bulletin* 24 (1998), p. 563.

15 Psychologists are forced by their conceptualizations to understand punitive judgements as effects caused by the occurrence of anger. One may however see them as part of the same process as anger. Thesis 4 is cashed out in terms that are compatible with both views.

The link between anger occurrences, our sense of justice and punitive attitudes has also been studied in connection with the vast literature on the so-called Ultimatum Game.[16] In the standard version of this game, one individual (proposer) controls an amount of money (say $10) and makes an offer to another individual (responder) on how to divide the $10 between the two individuals. Both individuals know the amount being divided and the rules of the bargaining. The responder can either accept or reject the offer. If the offer is accepted, the sum of money is divided as proposed and the bargaining ends. If the offer is rejected, both individuals receive nothing and the bargaining ends. As Srivastava *et al.* explain:

> The game–theoretic, sub-game perfect equilibrium, prediction is that a proposer should offer the smallest unit of currency and the responder should accept. The rationale is that an income maximizing individual would accept any offer since something is better than nothing. In contrast to the normative prediction, two robust findings have emerged in the literature (Camerer & Thaler, 1995; Guüh, 1995). First, proposers typically offer about 30–40% of the total amount, with a 50–50 split often the mode. Second, responders typically reject offers that represent less than 25% of the total amount. These findings suggest that individuals' behaviour is not entirely driven by self-interest. … The finding that responders are more likely to accept small offers when they come from a random device than from a human agent suggests that individuals punish unfairness and are not merely rejecting inequality (Blount, 1995). The willingness to sacrifice one's own interests (i.e., at a cost to one self) to punish those who are being

[16] The literature makes a reference to our feeling of fairness rather than our sense of justice. In this context, however, we take this distinction to be irrelevant.

> unfair suggests that emotions may underlie responders' rejection decisions.[17]

Note also how the nature of the game is such as to fall naturally in line with the idea that punitive attitudes are retributive. Players are aware that the traditional Ultimatum Game is a one-off interaction. Any impulse to punish at a cost to oneself, then, cannot be justified in terms of the effect that it may have on the proposer in future interaction. What is more, in their study, Srivastava *et al.* examine the extent to which in the Ultimatum Game, the cognitive appraisal of unfairness leads to the emotion of anger, which in turn, drives punitive behaviour (i.e., the rejection of offers).[18] The evidence gathered by the authors indeed suggests that anger mediates the influence of offer size on rejection rates as well as the influence of unfairness appraisals on rejection rates, evidence once again in line with Theses 4 and 5.[19]

[17] J. Srivastava, F. Espinoza, A. Fedorikhin, "Coupling and Decoupling of Unfairness and Anger in Ultimatum Bargaining," *Journal of Behavioral Decision Making* 22 (2009), pp. 475–489, on p. 476.

[18] *Ibid.*

[19] *Ibid.*, p. 481. The mediating role of anger was further confirmed in two ingenious ways. First, the authors decoupled the cognitive appraisal of unfairness from the anger reaction. This method is relevant against the background of research showing that behavioral response driven by emotions, and anger in particular, can be altered by leading people to believe that the emotion being experienced is caused by an external, unrelated source. Strivasana *et al.* hence induced their subjects to believe that their anger was to be attributed to something other than the unfair offer. They then recorded that rejection rates in this group fell to 60% as compared to 93% in the control group (*Ibid.*, p. 483). Secondly, in their final study (*Ibid.*, pp. 484-486), the authors confirm that the effect is explained by the specific emotion of anger as opposed to negative valenced emotions in general. Note finally that studies using other bargaining tasks (G. Ben-Shakhar, G. Bornstein, A. Hopfensitz, F. van Winden "Reciprocity and emotions in bargaining using physiological and

As Wiegman notes,[20] studies such as Lerner *et al.*[21] and Strivastava *et al.*[22] deal purely with incidental anger, i.e., how the anger caused by one person has spill-over effects on punitive attitudes directed to other persons. The evidence gathered by Fabiansson and Denson remedies this shortcoming thereby providing further evidence in favour of Theses 4 and 5.[23] Participants gave a brief speech about their life goals to a fictitious participant and were subsequently either insulted or not by the fictitious participant. Next, participants played two ultimatum games against the fictitious participant and non-provoking control counterparts. Across two economic bargaining tasks the authors found that provoked participants punished the speech task counterpart more than unprovoked participants. Angered participants were more likely to give money to a novel participant than the person who provoked them. Angered participants also proposed less fair offers to the speech task counterpart than participants who were not provoked. Finally, they were less willing to accept offers from the speech task counterpart regardless of how fair the offer was.

Beside evidence to the effect that anger incidentally and directly modulates punitive responses in situations in which one is

self-report measures," *Journal of Economic Psychology* 28 (2007), pp. 314–323) have also shown that physiological arousal and self-reported anger are associated with punishment decisions.

20 I.T. Wiegman, *Anger and Puishment: Natural History and Normative Significance* (Ph.D. Dissertation, Washington University in St. Louis 2014), p.14.

21 J.S. Lerner, J.H. Goldberg, P.E. Tetlock, "Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of responsibility,"

22 J. Srivastava, F. Espinoza, A. Fedorikhin, "Coupling and Decoupling of Unfairness and Anger in Ultimatum Bargaining."

23 E.C. Fabiansson, T.F. Denson, "The Effects of Intrapersonal Anger and Its Regulation in Economic Bargaining," PLoS ONE 7 (2012), https://doi.org/10.1371/journal.pone.0051595

personally responding to unfair treatment, there is also evidence that anger modulates our responses in cases of so called altruistic or third-party punishment. The latter is the type of behaviour displayed by those that incur costs in order to punish someone who has not directly harmed the subject or those one cares about. Third-party punishment is quite important to our present concerns, as it may be taken to model criminal punishment. Nelissen and Zeelenberg manipulated anger and guilt in a Dictator Game, a bargaining game similar to the Ultimatum Game in which, however, the proposer dictates the division of the money and the responder or receiver has no say.[24] In the particular setup of their study, the opportunity to punish the allocator was given to a third-party as opposed to the receiver (who had this opportunity in the Ultimatum Game). According to the authors, the evidence gathered suggests that anger and guilt independently constitute sufficient but not necessary causes of punishment. Low levels of punishment are observed only when neither emotion is elicited. As Nelissen and Zeelenberg note in their general discussion,[25] the impact of anger demonstrated in their studies is in line with views that hold punishment primarily to serve retributive purposes.[26] Theses 4 and 5 seem to receive support from these studies.

Importantly, however, this evidence may seem in contrast with Batson *et al.*, whose results indicate that anger is reported by subjects only when either subjects were directly harmed by unfair

[24] R.M.A. Nelissen and M. Zeelenberg, "Moral emotions as determinants of third-party punishment: Anger, guilt, and the functions of altruistic sanctions," *Judgment and Decision Making* 4 (2009), pp. 543-553.
[25] *Ibid.*, pp.548-549.
[26] K.M. Carlsmith, J.M. Darley, P.H. Robinson, "Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment"; J.M. Darley & T.S. Pittman "The Psychology of Compensatory and Distributive Justice," *Personality and Social Psychology Review* 7 (2003), pp. 324-336.

treatment (personal anger) or someone whom the subjects care about or identify with was harmed by unfair treatment (empathic anger).[27] However, no anger was reported by subjects who observed unfair treatment being imparted to someone other than themselves or someone whom they cared about, i.e., anger at the sheer violation of a moral norm of equity or fairness (moral outrage). Batson *et al.*, however, are mute with regard to punishing behaviour. In particular, it does not probe whether those who observe moral violations in the absence of personal or empathic anger, would still typically tend to want to punish the perpetrator. If they did, then we would have some evidence that clearly goes against some of the evidence discussed above. If they did not, however, we may have to consider the possibility that the scope of personal and empathic anger is wider than hitherto thought, so as to actually cover violations that experimenters have intuitively considered as impersonal (as the ones in the dictator game).

All in all, the evidence available in psychology indicates not only that anger and punitive attitudes co-vary but that the former plays a causal role in bringing about and modulating the latter. What is more, most of the studies do focus on bargaining games that seem particularly fitting in showing the retributive nature of punitive attitudes. As argued above, this is evidence in favour of Theses 4 and 5.

The type of studies discussed so far focuses exclusively on the connection between *occurrences* of anger and *occurrences* of punishment judgements. This methodology excludes by fiat a number of potential connections between anger or other justice-

[27] C.D. Batson, C.L. Kennedy, L-A Nord, E.L. Stocks, D.A. Fleming, C.M. Marzette, D.A. Lishner, R.E. Hayes, L.M. Kolchinsky and T. Zerger, "Anger at Unfairness: Is it Moral Outrage?," *European Journal of Social Psychology* 37 (2007), pp. 1272-1285.

related emotions and punishment. For one, it is blind to the possibility that anger may play a causal role in shaping our punitive judgements in other ways, as for example, the way suggested by Thesis 3. Fortunately, however, Milburn *et al.* present evidence to the effect that children who underwent harsh punishment are more likely to endorse harsher forms of punishment as adults such as capital punishment, and, in line with Thesis 3, this effect is mediated by the tendency of these individuals to experience anger ("trait anger").[28]

The Emotive Account of punishment would of course be strengthened by evidence showing, for example, that individuals whose capacity for justice-related emotions was damaged earlier on in their development do not display punitive attitudes or that cultures or genders that are less prone to justice-related emotions tend to display comparatively less punitive attitudes. Even in the absence of such evidence, however, we shall take the above to be sufficient evidence to hold on to the Emotive Account. As argued above, however, while our practice of punishment is shaped by our justice-related emotions, this is not to say that we are inclined to punish and consider punishment deserved exclusively when *in their grip*. These emotions shape individuals in their development and social interactions and have shaped human institutions throughout the ages. We may well have grown to understand when punishment is deserved even in the absence of occurrences of justice-related emotions just as we may recognize while in the grip of anger that punishment would be undeserved. These normative issues will occupy us next.

[28] M.A. Milburn, N.M. Niwa and M.D. Patterson, "Authoritarianism, Anger, and Hostile Attribution Bias: A Test of Affect Displacement," *Political Psychology* 35 (2014).

V

The Internal Logic of Punishment

Proving a tight connection between punitive attitudes and justice-related emotions has an important advantage: just as our emotional reactions are susceptible of normative assessment, so can the punitive reactions that are a part of it. In fact, the latter inherit the normativity of the former, or so we will argue in this section.

Emotions, and anger as one of them, have *formal objects*.[29] An emotion's formal object plays a double role. First, it is essential in making each emotion intelligible as the type of emotion it is. Thus, for example, *danger* is thought to be the formal object of fear: fear is the apprehension of a particular object as dangerous or frightful.[30] The second role of formal objects has to do with the normativity of emotions. Emotions are not simply taken to be brute affective reactions but are assessable as more or less appropriate, fitting, or rational. Formal objects afford the norm against which this assessment is conducted, or again, the correctness conditions for emotional occurrences. Thus, for example, a particular object that is not extremely dangerous makes extreme fright a disproportionate and hence inappropriate (irrational, unfitting, or incorrect) emotional response.

[29] See Teroni (2008) for a comprehensive and up to date discussion of emotions and their formal objects.

[30] The apprehension at hand may happen at the personal or the sub-personal level and mastery of the concept of danger or frightfulness is not necessary to experience fear. Note also that emotions are not identified solely in terms of their formal objects. When characterizing an emotion, we also typically appeal to its phenomenology, the way it feels, and to its action tendencies, i.e., what it typically disposes one to do. Hence, for example, our experience while in the grip of fear will feel quite different from our experience while in the grip of guilt. Similarly, while the former emotion will typically dispose us to flee, the other will typically dispose us to make amends.

Importantly, the normative assessment of emotions can focus on distinct dimensions. In particular, we can distinguish normative assessments that are internal to the emotion from those that are external to it. On the one hand, internal normative assessments are judgments about the fittingness or appropriateness of an emotion to its object such the ones just described. They are internal in the sense that what regulates them, i.e., the formal object, is the very same thing that serves to make each emotion intelligible as the emotion it is. Hence, for example, judgments about the appropriateness of fear are regulated by danger, which in turn is what we use to understand the emotional occurrence at hand as an occurrence of fear. Hence, whether your fear of this spider is appropriate will, in this internal sense, be regulated by the danger that the spider poses.

External normative assessments, on the other hand, do not use the emotion's formal object as their norm. Hence, even though fear may be appropriate if a large bull were charging, it may on that occasion be prudentially best not to feel fear (if one could) and not to do whatever fear disposes one to do (typically, flee). Similarly, even if it were fitting to feel envy towards your rival because he has something good that you lack, it may well be that, from a moral point of view, it is never good to feel envy.[31]

Let us move on to the justice-related emotions in general and anger in particular. It appears that the formal object of anger is an injustice most often in the shape of an unjustified culpable wrongdoing. In anger, that is, we typically apprehend someone as the culprit of an unjustified wrong. On the basis of this, we typically experience feelings of injustice, cognitions to the effect that the wrongdoer deserves a disvalue for what he has done, and punitive action tendencies accompanied by feelings of entitlement

[31] See J. D'arms and D. Jacobson, "The Moralistic Fallacy," *Philosophy and Phenomenological Research* 61 (2000), pp. 65-90, for an elaboration of this point.

or righteousness. Anger is then internally regulated by unjustified culpable wrongdoing. If, for example, you incorrectly believe that Sam has stolen your bike, your cognition that he deserves punishment and any punitive action that you might initiate would be unjustified or inappropriate. Similarly, for punitive attitudes you may have towards someone who is incapable of responsible agency. Finally, extreme anger accompanied by harsh punitive attitudes for a minor unjustified culpable wrongdoing would be disproportionate, unfitting, or unreasonable. The wrongdoer would not deserve such harsh punishment.

In short, then, the normative logic internal to punishment is similar to that of anger and other justice-related emotions. Being essentially connected to these emotions, internally, punishment is regulated by considerations of culpability and seriousness of wrongdoing. The internal logic of punishment is purely retributive. Just as any other emotion, however, anger, for example, and its punitive action tendencies may also be normatively assessed externally via considerations such as expediency, deterrence, rehabilitation, education, rights, and morality. Hence, while internally, questions about punishment are purely retributive, externally, such questions are regulated by a potentially open and diverse set of considerations distinct from considerations of deservingness. Hence, for example, while in the grip of an appropriate occurrence of anger, you may be correct in thinking that someone deserves punishment and in wanting to inflict it upon this person. Yet there may be external considerations such as prudence to the effect that punishment on this occasion (or on all occasions relevantly similar to this one) is not what you have most reason to do (it may be too costly or even dangerous to punish the person in question).

Let us sum up our line of argument so far. We started off with the idea of *legal* punishment and then transitioned to that of *generic*

punishment in order to shed some light on the former. In the process, we have learnt that punishment has a retributive core, which it inherits from the justice-related emotions on which it is based. More in detail, we seem to have reached four significant conclusions:

I. Something counts as punishment only if it is a disvalue inflicted as an intentional response to a perceived wrongdoing.

II. Punishment is always either assessed as deserved to some degree or assessed as undeserved.

III. Judgements about the deservingness of punishment are internally regulated by the formal object of justice-related emotions in terms of culpability and seriousness of wrongdoing, that is, purely in retributive terms.

IV. Normative judgements about punishment are also regulated externally by an open and potentially diverse set of considerations (e.g., education, deterrence, etc.) that are distinct from deservingness.

In line with what claimed in section 1, we take the four theses above to afford a psychologically based, interpretative account of our punitive practices. We have called this account the *Emotive Account* of punishment. With these conclusions in hand, it is time to retrace our steps back to legal punishment and determine what normative consequences follow for it.

VI

Emotions, Punishment, and the Criminal Law

The first important consequence of the *Emotive Account* is that any account of the criminal law that purports to be built around the notion of punishment will carry with it the retributive logic that is internal to this practice and its emotive basis. As claimed above, in the absence of these, our ideas and practices of punishment would not be intelligible.

This is a momentous conclusion for the criminalization question and that from the perspective of both instrumentalists and non-instrumentalists. The latter will now have at least one clear substantive account of the distinctiveness of the criminal law. While this may sound as good news for non-instrumentalists, as we will presently argue, it may not be as good news for them as it at first appears to be. As for instrumentalists, they will now face the task of explaining how from their point of view we should be interested in a practice whose retributive core is at least on the face of it insensitive to considerations of an instrumentalist kind. Let's elaborate this last challenge first.

Consider, for example, an account such as Tadros' where an instrumentalist, non-consequentialist view for the justification of punishment is offered.[32] On this view, what distinguishes punishment from non-punitive penalties is the idea that the former primarily involves making the offender suffer, while the latter is supposed to ensure fairness in the distribution of resources. This distinction provides a basis to determine the scope of the criminal law, which, he claims, is and ought to be about punishment rather than penalties. An upshot of this instrumentalist view is that punishment is imposed on people as a

[32] V. Tadros, "Criminalisation and Regulation.," in R.A. Duff, L. Farmer, S.E. Marshall, M. Renzo (eds.), *The Boundaries of the Criminal Law*.

means to prevent further wrongdoing by others provided that the constraint on inflicting pain on people is lifted because of their wrongdoing.

If the view of punishment defended above is correct, an instrumentalist view of this kind is untenable. Punishment is not only the infliction of suffering, for that is also what assault amounts to. Neither will it do to mention that punishment can be inflicted provided that the offender's wrongdoing lifts the constraint not to inflict pain on him. This still fails to identify punishment. Punishment is rather a *response* to a perceived wrongdoing. To inflict pain, hard treatment, or undesirable experience as a deterrent for others, however, is not to inflict pain, hard treatment, or undesirable experience as a *response* to an offender for what he has done. Deterrence is not necessarily a response. The fact that punishment deters, if that were indeed a fact, is incidental to punishment.

What is more, an account such as this ignores the idea of deservingness. As argued, the concept of punishment comes with that of deservingness. The instrumentalist may reply that, insofar as her account is purely justificatory, the conceptual link between punishment and deservingness may be blissfully ignored. Whatever is true of the concept of punishment, only deterrence counts at the justificatory or normative level. The problem with this stance, however, is that it deeply violates the normative elements inherent to the concept and practice of punishment. If the account were correct, legal discourse about punishment deservingness would be either unjustified or only justified to the extent to which it served the aim of deterrence. This, however, is not the way in which we understand and use this concept. It would distort its meaning. Punishment deservingness follows a retributive logic that is often impervious to non-retributive

considerations. Its peculiar type of normativity imbues our discourse.

The instrumentalist may at this point want to dig her heels. Legal punishment, if justifiable, must come in line with deterrence or whatever other considerations instrumentalists take to provide the ultimate justification of punishment. Deservingness and its retributive logic are acceptable only to the extent to which they fall in line with the instrumentalist logic. Purely teleological views of the justification of punishment, for example, are common place in the relevant Italian and post-1966 (post Alternativ-Entwurf) German criminal jurisprudence.

The question that our analysis helps us pose, however, is the following. How far would these jurists continue to support their justificatory views if the latter became more and more at odds with the retributive logic of punishment, i.e., if there was an ever widening gap between what people took to be deserving and undeserving of punishment, on the one hand, and what the jurists took to be justifiable legal punishment, on the other? By making punishment just another tool for the regulation of behaviour, the risk that instrumentalists incur is to deprive punishment of its very nature and make its legal practice unrecognizable and alienating. In this respect, instrumentalists seem to lack the correct understanding of the human practice of punishment.

The *Emotive Account* is therefore inimical to instrumentalist positions about criminalization and the justification of punishment. This may be thought to be good news for non-instrumentalism, at least insofar as it is in line with the retributivist features of the account. The news, however, is not so good for non-instrumentalism either, though for different reasons.

Let us consider here a non-instrumentalist view such as Duff, which defends a partly retributivist partly teleological view of the justification of punishment.[33] On this view, the retributive logic of punishment is given heed. Such non-instrumentalist view will also be able to show that legal punishment is more than just another instrument for the regulation of behaviour in the hands of the state in that its nature, or at least its justification, is quite unlike that of non-punitive sanctions. However, in deciding what and how to punish, non-instrumentalism is bound to accept that the state has aims other than retribution. These are the aims and constraints mentioned above, whose justificatory force is external to the retributive logic of punishment. Liberal democracies typically appeal to deterrence and rehabilitation as aims that criminal punishment should secure. Criminal policies and decisions, then, must be in line with these aims and constraints in order to be justified.

But in the context of liberal democracies, these aims are only the tip of the justificatory iceberg. Deterrence, for example, has any weight only insofar as a state aims at ensuring the safety (or security) of its citizens, where safety is in turn valuable insofar as it contributes to citizens' well-being or, perhaps, as it is necessary to respecting their rights and autonomy. Many in fact believe that liberal democracies derive their ultimate justification from the ideal of their citizens' equality and autonomy, where the latter is minimally understood as involving a kernel of liberal rights including the right of citizens to participate in government and the freedom to decide how to lead one's life compatible with the equal freedom of all other citizens.

Importantly, however, if this justificatory story is correct, the ideals of equality and autonomy will regulate not only the external

[33] R.A. Duff, *Punishment, Communication, and Community*.

logic of punishment but also its internal one. Or more precisely, given that the internal logic of punishment cannot be internally regulated by anything other than itself, given their ultimate justificatory status, these ideals will externally regulate punishment. This means that retributive considerations of deservingness must either fall in line with such ideals or be systematically overridden or excluded by them. The question, then, is whether retributive punishment can fall in line with such ideals and, if so, to what extent.

In the recent debate, Dubber is one of the few authors that attempts a reconciliation of autonomy as the ultimate source of legitimacy for the liberal state, on the one hand, with the criminal law understood as centrally involving retributive punishment, on the other.[34] According to Dubber and the democratic republican tradition to which he appeals, to say that autonomy or self-government is the ultimate source of state legitimacy is to say that the governed must, directly or indirectly, consent to its actions. "This means that, to put it bluntly, punishment in a democratic republic can be legitimate only as *self-punishment*."[35] Dubber then goes on to examine how the ideal of autonomy understood in this way can be made to square with, respectively, the definition of criminal laws (the realm of criminal law), their application to a particular case (the realm of criminal procedure law), and the infliction of sanctions (the realm of prison law).

For our purposes, it is most important to focus on the definition of criminal laws or the realm of criminal law. In short, on the view proposed by Dubber citizens have, *qua* persons, a right to their autonomy; the function of the law is to protect

[34] M.D. Dubber, "A Political Theory of Criminal Law: Autonomy and the Legitimacy of State Punishment" (March 15, 2004). Available at SSRN: http://ssrn.com/abstract=529522 or http://dx.doi.org/10.2139/ssrn.529522.
[35] *Ibid.*, p. 5.

autonomy;[36] and *criminal* law helps the state discharge this function through punishment. Crime is defined as an autonomous attempt on behalf of a person to compromise or destroy another person's autonomy. When crime has occurred, a person's right to her autonomy has been violated. At this point, punishment becomes the "vindication" or "the dramatic reaffirmation" of the victim's autonomy, as it communicates to the world that the offender's attempt to deny the victim's personhood was unsuccessful.[37] Finally, not only does the victim have the right to have the offender punished, but the offender himself has the right to be punished, insofar as treating him as an ahuman source of danger denies him the "dignity and respect" he "deserves" as a person.

This view contains many interesting claims, such as, for example, the claim that all crime should be conceived as a violation of autonomy[38] or that offenders have a right to be punished.[39] Given our purposes, however, the view's main difficulty consists in explaining how exactly inflicting pain, hard treatment, or undesirable experience on the offender as a response to her action will *vindicate* the victim's *autonomy*, in particular when the offender is himself understood as having rights to autonomy. Why would sending the offender on a luxurious cruise as opposed to inflicting pain, hard treatment, or undesirable experience on him not vindicate the victim's autonomy? What is it about infliction of pain, hard treatment, or

[36] *Ibid*., pp. 30-33.
[37] *Ibid*., p. 33.
[38] See J. Stanton-Ife, "Horrific Crime," In R.A. Duff, L. Farmer, S.E. Marshall, M. Renzo and V. Tadros (eds.), *The Boundaries of the Criminal Law*, for an argument to the opposite conclusion.
[39] Deigh doubts this view: J. Deigh, "On the Right to Be Punished: Some Doubts," *Ethics* 94 (1984), pp. 191-211.

undesirable experience that makes this a particularly appropriate way to vindicate violated autonomy?

Even if retributive infliction of pain, hard treatment, or undesirable experience could be shown to be particularly appropriate from the victim's point of view, how does it square with the view that offenders too are autonomous beings? While it is clear that non-punitive coercion can coherently be envisaged as facilitating compliance with a system of rules aimed at realizing liberal ideals, on a par with more positive incentives such as economic ones, it is far from clear that coercion of the kind involved in retributive punishment would be similarly suited. Retributive punishment does not as such aim at giving citizens-as-rule-followers reasons to comply with rules. Rather it focuses exclusively backwards, on citizens-as-rule-violators, on what they deserve in light of what they have done.

None of this is to say that we find punishing wrongdoers unfitting or, in fact, that we should do so. The point or points are rather these. First, whatever plausibility we find in the idea and practice of punishing wrongdoers is there prior to any story about victims' and offenders' autonomy. The "logic" inherent to liberal autonomy cannot explain, take over, or recast in its own terms the internal logic of punishment. Second, if anything, liberal autonomy, at least as discussed above, is inconsistent with such logic.

While we cannot exclude the possibility of non-instrumentalist views whose liberal credentials better conform with the internal logic of punishment, we hope the above to provide enough material for at least initial scepticism with regard to both instrumentalism and (liberal) non-instrumentalism.

VII

Conclusion

The argument above presents considerable difficulties for both instrumentalism and non-instrumentalism about the criminal law. While the former is misguided about the distinctive nature of the criminal law and runs the risk of developing a practice that is potentially detached from its human understanding, the latter correctly describes the criminal law as distinctive tool for the regulation of behaviour but seems, at least initially, at odds with the justificatory background of liberal states. This, of course, is not to say that normative theorists should give up developing views of this kind. Yet, it is a conclusion that does incline us to consider what alternative views of criminal punishment there are. In line with one part of our conclusion, one option is to accept that there is an unsurmountable tension between retributive punishment and liberal ideals. We would then have to understand whether it matters more to us to maintain such ideals while abandoning retributive practices or to maintain our retributive practices while abandoning or relaxing certain liberal ideals. Yet, in line with the other part of our conclusion, it may simply not be an option to change our retributive practices by simply altering their normative logic from the outside. These practices ultimately rely on deep engrained mechanisms and meanings. While not impossible, their modification or abandonment should be expected to be a long and difficult process.[40]

Aarhus University

[40] We are grateful to Johanna Seibt and Alessandro Spena for their incisive comments.

If you need to cite this article, please use the following format:

Katrine Krause-Jensen & Rodogno, Raffaele, "Criminal Law and the Internal Logic of Punishment," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), 135-171, edited by S. Maffettone, G. Pellegrino and M. Bocchiola

SYMPOSIUM

THE PHILOSOPHY OF PUNISHMENT

THE REACTIVE SENTIMENTS AND THE JUSTIFICATION OF PUNISHMENT

BY

ANDREW ENGEN

© 2015 – *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015): 173-205
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

[THIS PAGE INTENTIONALLY LEFT BLANK]

The Reactive Sentiments and the Justification of Punishment

Andrew Engen

The conviction that those who commit serious acts of wrongdoing should be punished is secure in our moral thinking. Public debate about whether an accused murderer or rapist should be punished focuses on his guilt or innocence, not whether he should be punished if he is guilty.[1] Certain modes of punishment, such as the death penalty, are sometimes the subject of vigorous debate, but the idea that murderers and rapists should be punished is not. A theory of punishment that takes commonsense morality seriously must give an account of the positive reason the state has to punish those who commit serious crimes in our society. Without such a reason, the state would be devoting its limited resources to pointlessly harming its citizens. I will call this the *positive reason* desideratum. In order to make sense of our moral thinking, this account should be sufficiently general: it should explain why we always, or almost always, have positive reason to punish serious crimes.

It is remarkable that the idea that serious wrongdoers should be punished is so entrenched in our moral thinking. For no less entrenched is a seemingly conflicting moral conviction: that

[1] This is not simply out of respect for the rule of law. The applications of laws that are controversial are themselves often the subject of public debate, in many cases to highlight the injustice of those laws.

Philosophy and Public Issues (New Series), Vol. 5, No. 1 (2015): 173-205
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

people have moral rights to goods such as liberty and property. Our conventional modes of punishment, such as imprisonment, seem to violate this set of rights. Moral philosophers writing on punishment have largely been concerned with this issue—not why we have positive reason to punish those who commit serious crimes, but how we are ever permitted to punish any persons who normally bear rights to liberty and property. A theory of punishment must explain why the punished cannot reasonably object to their punishment, even though it deprives them of goods to which they otherwise have a right. I will call this the *no valid objection from rights* desideratum.

In this paper, I defend an expressive justification of punishment grounded in our emotional responses to wrongdoing. This "reactive theory" is constructed out of a Strawsonian account of moral responsibility that understands our practice of holding one another responsible in terms of the reactive sentiments of resentment, indignation, and guilt. Punishment is morally justified on the theory insofar as it appropriately expresses the justified indignation of the community in response to serious wrongs. I contend that this blaming function of punishment avoids some familiar problems encountered by traditional justifications of punishment in satisfying the *positive reason* and *no objection from rights* desiderata, which makes it a promising centerpiece of a theory of punishment.[2]

[2] In trying to make sense of "our" moral thinking about punishment, I will be drawing upon my own intuitions as well as my understanding of the practices of moral responsibility and the system of punishment in the liberal democratic, Western society in which I live. This bias is apparent in my formulation of the *positive reason* desideratum. The notion that "we" have reason to punish addresses readers who have some access to political power. Because the reactive theory justifies punishment as a reaction of the community to crime, it will only apply to those societies where the state acts on behalf of the people.

I

Deterrence and Retribution

One traditional strategy for justifying punishment focuses on its on its deterrent function. According to deterrence theories, punishment is justified by the role that it plays in reducing crime. Imprisonment, for example, both provides a disincentive for potential lawbreakers and incapacitates those who have shown a disposition not to respect the law. Deterrence theories have obvious appeal: the reduction of serious wrongdoing is clearly an important social goal. Yet critics question whether the deterrent function can satisfy each of the *positive reason* and the *no objection from rights* desiderata.

We can conceive of scenarios in which punishing a serious criminal would not deter crime. Consider the assassin of a civil rights leader in the distant past. He is not apprehended until years after the murder. Great social progress has been made so he and others are no longer disposed to commit similar crimes because the political ends they were intended to achieve are no longer achievable. We can suppose that punishing the assassin would do nothing to reduce crime. The deterrent function of punishment does not explain why, intuitively, we still have positive reason to punish the assassin for his atrocious wrong. In response to this counterexample, it might be argued that a deterrence theory could come up with *some* explanation of how punishing him would deter crime, given more information about the case. Such a strategy cannot, however, make sense of our moral thinking; our shared conviction that we have positive reason to punish the assassin does not depend on such information.

The most common objection to deterrence theories casts doubt on whether they can satisfy the *no valid objection from rights* desideratum. The objection holds that the state is not justified in

punishing a person known to be innocent, even if doing so would effectively deter crime. Intuitively, an innocent person has a valid objection to her imprisonment with appeal to her rights, regardless of the deterrent effect of her imprisonment. A deterrence theorist could claim that the bad consequences of punishing the innocent never actually outweigh the good.[3] But the validity of an innocent person's appeal to her rights to liberty or property do not seem to hang in the balance as we weigh such considerations. This points to a larger problem with the consequentialist character of deterrence justifications: they treat the punished as causal levers in a system that aims to make society secure.[4] In commonsense morality, rights are thought to limit how people may treat one another independently of the consequences of that treatment. The state is not normally justified in violating people's rights simply on the basis that doing so would lead to good social consequences. Deterrence considerations alone do not explain why such treatment is justified when we punish serious crimes.

The traditional rival to deterrence theory, retributivism, is based on the idea that wrongdoers deserve to suffer in proportion to the harm they have caused their victims. Punishment fulfills its retributive function by inflicting this deserved suffering on the punished. Serious criminals do not have a valid objection to the deprivations of punishment because those deprivations ensure the criminal gets his just deserts. Retributivism justifies punishment as a backward-looking *response* to serious wrongs,

[3] For a well-developed version of the criticism that deterrence theory justifies punishment of the innocent and a utilitarian response along these lines, see H. J. McCloskey, "A Non-Utilitarian Approach to Punishment" *Inquiry* 8 (1965): 239-63 and T. L. S. Sprigge, "A Utilitarian Reply to Dr. McCloskey" *Inquiry* 8 (1965): 264-91.

[4] For a prominent presentation of this criticism, see Jeffrie Murphy, "Marxism and Retribution," *Philosophy and Public Affairs* 2 (1973): 217-43.

rather than as a forward-looking device for achieving desirable consequences and thereby avoids some of the difficulties that plague deterrence theory. Unlike its traditional rival, retributivism does not take an instrumental view of criminals, but instead responds to them as moral agents.

Nevertheless, there is reason to be skeptical of the retributive idea that grounds this justification of punishment because it is in tension with the rest of our moral commitments. We generally think it is morally inappropriate to hold anyone's suffering to be intrinsically good.[5] The idea that wrongdoers deserve to suffer in proportion to the harm they have caused is difficult to square with this moral commitment. Upon reflection retributivism can seem to be, in the evocative phrase of H. L. A. Hart, "a mysterious piece of moral alchemy in which the two evils of moral wickedness and suffering are transmuted into good."[6] Reasonable skepticism about the retributive idea calls into question retributivism's claim to satisfy both the *positive reason* and the *no objection from rights* desiderata. While a convincing argument for the retributive idea could defuse this skepticism, attempts to ground the retributive idea are notoriously obscure.[7] With respect

[5] See, for example, A. C. Ewing, *The Morality of Punishment* (London: Kegan Paul, Trench, Trubner & Co., Ltd, 1929), 29: "In every other instance the deliberate infliction of pain is wrong, except where necessary as a means to happiness or ethical improvement, in every other instance our primary duty is to abstain from bringing evil on our fellow-men…. Yet here we are asked to inflict pain for pain's sake. It seems strange that a kind of action which under ordinary conditions is regarded as the very extreme of moral depravity should become a virtue in the case of punishment."

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

[7] Both critics and defenders of retributivism emphasize this point. See, for example, Ted Honderich, *Punishment: The Supposed Justifications Revisited* (London: Pluto Publishing, 2005), 196; George Sher, *Desert* (Princeton: Princeton University Press, 1987), 69.

to the first desideratum, such skepticism undermines the view that the state should spend its resources inflicting punishment on serious criminals so that they get their deserved suffering. With respect to the second, a criminal could reasonably object to the deprivations of punishment if those deprivations are justified in their role of inflicting the suffering he purportedly deserves.

Even if we assume that wrongdoers deserve to suffer, retributivism faces a further difficulty satisfying the *positive reason* desideratum. Intuitively, it seems that the state has reason to punish even when punishment does not inflict suffering on the punished. Consider less severe forms of punishment, such as community service. We do not think that the punishment of someone who finds that he enjoys community service fails to serve its purpose because it does not make him suffer. Of course, we do not punish the most serious criminals, such as murderers, with community service. However, I do not believe that this is because other forms of punishment, such as imprisonment, carry out the retributive function. I grant that the life of someone in prison often involves a great deal of suffering; confinement in small spaces often takes a physical and an emotional toll. But imprisonment does not always inflict more suffering on inmates than they would have experienced on the outside. Sometimes imprisonment takes people out of dire and dangerous situations and provides them with food, shelter, and medical care that they otherwise might not have had. When punishing someone causes less suffering in a person's life than not punishing him, it does not carry out its purported retributive function.[8] Yet

[8] Some retributivists might claim that this objection reveals that I misconstrue what the retributive idea is. It is not that wrongdoers deserve to be worse off overall because of their wrongdoing, but rather than they deserve to be deprived of some objective good in proportion to what they have deprived their victims, who, after all, may not have been made worse off overall by the

commonsense morality does not hold that such punishment thereby fails to be a fitting response to wrongdoing.

I have posed challenges to both the deterrent function and the purported retributive function with respect to fundamental features of our commonsense thinking about the morality of punishment. My argument has been brief and has not established that more sophisticated theories built around these respective functions could never meet the challenges. Instead of investigating those possibilities, however, I will go on to argue that the blaming function central to the reactive theory satisfies both the *positive reason* and *no objection from rights* desiderata in a relatively straightforward way. I believe that this provides reason to think that the theory better reflects our moral thinking about punishment than the traditional alternatives.

A defender of the traditional views might claim that my above argument ignores the possibility that traditional views could be combined in a way that avoids the objections I raised against each separately. General deterrence could provide the reason we have to set up a system of punishment, while the pursuit of this aim could be constrained by retributive considerations. Only those who had committed crimes would be subject to punishment because only they deserve to suffer, but they would be punished for the sake of the social good. Such a theory would evade the worries that I raised about the deterrence function's satisfaction

crime. Imprisonment deprives the imprisoned of important objective goods regardless of whether their lives are worse because of their imprisonment, thus it is a fitting response to wrongdoing. If that is the correct understanding of "retributivism," the reactive theory could be interpreted as a version of expressive retributivism that explains why it is fitting that wrongdoers are deprived some objective good. Though, as will become clear in section 3, the reactive theory understands proportionality differently than traditional retributivism.

of the *no valid objection from rights* desideratum and the retributive function's satisfaction of the *positive reason* desideratum. Nevertheless, the theory would still need to explain why we have positive reason to punish criminals when no deterrence would be achieved by doing so and why reasonable skepticism about the retributive idea does not ground a valid objection to punishment.[9] I believe that the reactive theory is superior to such a combined theory because it avoids these difficulties while capturing some of what makes the traditional views appealing. On the one hand, part of the explanation of why we are willing to contribute significant state resources to systems of punishment is undoubtedly that we think crime reduction is an important goal. I will contend that the nature of blame gives us reason to express it in ways that contribute to that goal. On the other hand, one reason why people are sometimes reluctant to give up on retributivism, even if they are skeptical of the retributive idea, is that retributivism acknowledges the agential status of the punished. The reactive theory offers an alternative, according to which we can treat the punished as responsible agents without being retributivists.

II

The Moral Importance of Blame

In his seminal 1965 article, "The Expressive Function of Punishment," Joel Feinberg criticizes contemporary philosophical

[9] Igor Primoratz emphasizes the first point in objecting to a two-level theory of this kind. Primoratz, *Justifying Legal Punishment* (New York: Humanity Books, 1989), 142.

discussions of punishment for ignoring the way in which punishment expresses both the community's emotions and also the judgment that the criminal acted wrongly.[10] He points out a number of important social roles that punishment plays in virtue of its expressive function.[11] However, he questions whether the harms characteristic of punishment are actually required to carry out that function.[12] A number of philosophers subsequent to Feinberg have offered justifications of punishment grounded in its expressive function.[13] The reactive theory fits in this tradition, focusing in particular on the emotions from the community that punishment expresses. According to the reactive theory, the state is justified in punishing someone insofar as punishment expresses well the community's appropriate indignation toward that person's crime. In this section, I will explain how systems of punishment can be conceived of as systems of blame. I will

[10] Joel Feinberg, "The Expressive Function of Punishment" reprinted in *Doing and Deserving* (Princeton, N. J.: Princeton University Press, 1970), 95-118.

[11] Ibid., 101-105.

[12] Ibid., 115-116: "One can imagine a public ritual, exploiting the most trustworthy devices of religion and mystery, music and drama, to express in the most solemn way the community's condemnation of a criminal for his dastardly deed. Such a ritual might condemn so very emphatically that there could be no doubt of its genuineness, thus rendering symbolically superfluous any further physical hard treatment. Such a device would preserve the condemnatory function of punishment while dispensing with its usual physical media—incarceration and corporal mistreatment. Perhaps this is only idle fantasy; or perhaps there is more to it. The question is surely open."

[13] See, for example R. A. Duff, *Punishment, Communication, and Community* (New York: Oxford University Press, 2000); Jean Hampton, "The Moral Education Theory of Punishment," *Philosophy and Public Affairs*, 13 (1984): 208-38; Uma Narayan, "Appropriate Responses and Preventive Benefits: Justifying Censure and Hard Treatment in Legal Punishment," *Oxford Journal of Legal Studies* 13 (1993): 166-82; Robert Nozick, *Philosophical Explanations* (Cambridge, Mass.: Harvard University Press, 1981), 363-397; Andrew von Hirsch, *Censure and Sanctions* (Oxford: Oxford University Press, 1993).

maintain that where political institutions are suitably democratic, the state is in a position to blame on behalf of its citizens and that such blame is morally important. A full defense of the theory, like a full defense of deterrence theory or retributivism, would take up issues of political legitimacy, such as the conditions under which the state has the authority to punish criminals. In this paper, I will set aside those issues for the most part. My more modest aim will be to explain how the blaming function that justifies punishment according to the reactive theory satisfies the *positive reason* and *no objection from rights* desiderata, and how characteristics of the emotions that constitute blame provide the theory with tools for explaining why hard treatment is required to carry out punishment's expressive function.

In "Freedom and Resentment," P.F. Strawson argues that by looking at our practice of holding one another responsible, we can see that ascriptions of moral responsibility would not be threatened by the truth of determinism.[14] I will not assess Strawson's argument as a response to incompatibilism about responsibility and determinism here, but will take from his famous lecture the account of moral responsibility and its connection to the emotions. According to Strawson, when we hold agents responsible we are subject to a range of emotions that are given to us with the structure of human relationships. We feel resentment when we are wronged and indignation on behalf of others when they are wronged. Toward some people, such as very young children and the severely mentally disabled, we are not subject to these attitudes, taking instead what he calls "the objective stance" toward them. When we suspend the reactive sentiments toward someone in this way, we do not hold him morally responsible. Strawson maintains that these emotions are

[14] P.F. Strawson, "Freedom and Resentment," reprinted in *Free Will*, ed. Gary Watson (New York: Oxford University Press, 1982), 59-80.

constitutive of our practice of holding one another responsible: "the making of [a moral] demand *is* the proneness to such attitudes."[15] Holding people morally responsible involves holding them to moral obligations, so that holding someone morally responsible is being prone to the reactive attitudes in one's interactions with him, should he violate a moral obligation: resentment when he violates a moral demand toward oneself, and indignation on behalf of third parties when he violates moral demands toward them.[16]

Our practice of blame is closely related to our practice of holding people morally responsible. In developing a Strawsonian account of moral responsibility, R. Jay Wallace spells out the connection:

> [T]o blame someone is to be subject to one of the reactive emotions in terms of which the stance of holding people responsible is essentially defined, and these emotions are *expressed* by the sanctioning behavior to which the stance of holding people responsible naturally inclines us.[17]

Wallace's reactive account of blame identifies the reactive sentiments as the common element of all forms that moral blame takes—from private, unexpressed blame to public censure. He argues that accounts of blame that ignore the reactive sentiments leave out the attitudinal quality of opprobrium that is characteristic of blame.[18] This quality sets blame apart from other

[15] Ibid., 77.

[16] Strawson includes a number of other emotions among the reactive sentiments, but the connection of these emotions to moral responsibility is not obvious. See R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press, 1994).

[17] Ibid., 52.

[18] Ibid., 81. See also R. Jay Wallace, "Dispassionate Opprobrium: On Blame and the Reactive Sentiments" in *Reasons and Recognition: Essays on the Philosophy of*

negative assessments, such as criticisms of people's arithmetic or hairstyles, which are not typically accompanied by the reactive sentiments. Though the reactive sentiments are marked by a quality of opprobrium, they are not constituted by the retributive judgment that their objects deserve to suffer. One can reflectively endorse the idea that expressing the reactive sentiments is an appropriate response to wrongdoing without endorsing the idea that the suffering of the targets of those expressions is intrinsically good.

We can conceive of punishment as blame, on the reactive account of blame, if we can understand it to express the reactive sentiments. According to the reactive theory of punishment, punishment is justified when it expresses well the appropriate indignation of the community. The sense of "community" is to be understood broadly to include all citizens under the jurisdiction of the system of punishment. It is thus a condition on the reactive justification that the law and judicial system be responsive to the democratic will of citizens. It is beyond the scope of this paper to articulate how that condition may be met, but where it is, acts of punishment can be understood as externalized expressions of the indignation of the community. The criminal law sets out moral demands on members of the society in which it applies. In societies where the criminal laws are enacted by procedures that are answerable to the will of the community, the moral demands enshrined in the law are normative expectations to which citizens hold one another. According to the reactive account of moral responsibility, this stance of holding one another responsible is to be understood in terms of the reactive sentiments. When citizens hold all their fellow citizens to the normative expectation of following the laws,

T. M. Scanlon, eds. R. Jay Wallace, Rahul Kumar, and Samuel Freeman (New York: Oxford University Press, 2011), 348–72.

they believe it would be appropriate to feel the reactive sentiments when their fellow citizens violate those laws. The criminal law also delineates the punishment for those who violate its demands. In democratic societies, acts of punishment that deprive their objects of important goods, such as liberty and property, can be conceived of as expressing the indignation the community has endorsed as appropriate in response to the crime.

However, the democratic endorsement of punishment in response to some action does not always suffice for that punishment to express appropriate indignation. Emotional reactions can sometimes be irrational. Indignation would be inappropriate in response to actions that do not violate moral obligations. Punishment of such actions would not be justified by the reactive theory even in cases where such punishment reflects the will of the community. There are also actions that could be classified as blameworthy violations of moral obligations, but whose criminalization would violate citizens' civil liberties. In some cases criminalization of these actions would be inappropriate because punishing minor wrongs would express a higher degree of indignation than is appropriate. Moreover, the blaming function of punishment provides reason to punish only when embedded in a larger theory of democratic legitimacy that articulates the limits on state power, including presumably why some wrong actions ought not be criminalized.

Where public blame is called for, the reactive theory can account for the importance of its expression because the reactive emotions are bound up with values at the center of morality. When we feel resentment, indignation, or guilt about episodes of wrongdoing, we show that we value those people who have been victimized by it in a particularly important respect. These emotions are partially constituted by the value judgment that the person who has been wronged is owed moral consideration and

so is properly protected by moral obligations. When we are emotionally exercised on behalf of the mistreated party this reveals that we care about them in a way that simply intellectually recognizing their value does not.[19] In addition to showing that we value those who are wronged as being owed moral consideration, when we feel the reactive sentiments, we show that we also take the moral obligation whose violation inspires the sentiment to be important. This is reinforced when the underlying pattern of emotional response is general, involving a comparative susceptibility to the reactive sentiments in all cases in which the obligation is violated: a pattern of this kind demonstrates that we value certain ways of people getting on with one another. When we are susceptible to the reactive sentiments with respect to a moral obligation, we reveal that we take the obligation to be an important standard that ought to structure human interactions, and that we care that human interactions are structured accordingly.[20]

The values that underlie these responses give us reason not only to blame, but also to express blame publicly in a way that

[19] The idea that the resentment is essentially tied to self-respect is a main theme of Jeffrie Murphy's work on forgiveness. See, for example, *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1988), 16.

[20] Wallace suggests that the reactive sentiments reflect the valuing of a sort of relationship that moral norms make possible. "Dispassionate Opprobrium: On Blame and the Reactive Sentiments" in *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, eds. R. Jay Wallace, Rahul Kumar, and Samuel Freeman (New York: Oxford University Press, 2011), 369: "To internalize a concern for morality… is to care about relating to people in the distinctive way that is constituted through compliance with basic moral requirements. But people who care about this form of relationship naturally tend to hold themselves and others to the moral norms that are constitutive of it, where this in turn involves a susceptibility to the distinctively reactive sentiments."

speaks on behalf of the members of the community.[21] In going in for a system of blame that expresses indignation on behalf of all victims of serious crimes, a society affirms that all of its members are owed moral consideration. Failure to offer protection in the establishment and enforcement of laws to some members of the community signals a lack of respect for them. Part of what is objectionable about a society that punishes crimes against oppressed groups at a lower rate than those that target other citizens is that it treats people in the oppressed groups as less valuable. When, on the other hand, a community aims to hold all of those who commit serious crimes to account, it shows that it takes seriously the idea that all of its citizens have basic rights, rights whose violation it would be appropriate to be emotionally exercised about. It also reinforces the importance of those moral demands enshrined in criminal law as appropriately structuring relations between members of the community. Societies have reason to blame publicly in order for members to demonstrate that they value their citizens and take certain moral standards to be inviolable.

We have reason to blame serious crime not only to speak on behalf of members of the community, but also in a sense to speak *to* them: we have reason to blame in a way that citizens understand as expressive of blame. Anyone in society with an interest in seeing that those who commit serious wrongs are held accountable by the community for their wrongdoing should be able to understand that they are. One important potential audience is the family and friends of victims of crime. These people have an interest in having the victim's value affirmed by

[21] Speaking on behalf of the entire community in rejecting wrongdoing is one of the social roles of punishment described by Joel Feinberg, "The Expressive Function of Punishment" in *Doing and Deserving* (Princeton, N. J.: Princeton University Press, 1970), 103.

the community.[22] Marginal groups in society are another important audience. Members of these groups have a special interest in seeing the affirmation of those members of their groups who are crime victims. Marginalized members of society often have reasonable concerns about whether the institutions of society treat them as equals. By making good faith efforts to blame all those who commit serious crimes in ways that everyone in society can recognize as blame, a society can show that it takes all its citizens to be owed the moral consideration that makes blaming responses on their behalf appropriate.

According to the reactive theory of punishment, this public blaming function is the basis of the justification of punishment. Punishment is justified because it expresses indignation that appropriately blames criminals for serious wrongdoing. By subjecting criminals to punishment, we publicly acknowledge the moral standing of their victims, and show that we are committed as a society to the importance of certain moral standards in regulating our interactions with one another.

III

Why Not Say it with Weeds?

[22] The way that punishment affirms the sense of victims of crime as being wronged is highlighted by T.M. Scanlon, "Punishment and the Rule of Law" reprinted in *The Difficulty of Tolerance: Essays in Political Philosophy* (Cambridge: Cambridge University Press, 2003), 219-233. Scanlon, however, does not think that the expressive role of punishment justifies its hard treatment for a reason that I address in the next section.

I have argued that we have moral reason to publicly blame those who commit serious crimes because of the importance of the values to which blame is connected. Blame can take many forms, however, so why should we express it with punishment? If the indignation of the community could be expressed just as well (or better) without depriving criminals of their liberty or property, the blaming role of punishment would not justify those deprivations. In this section, I will argue that we have moral reason to express indignation toward serious crimes through depriving criminals of such goods.

Many philosophers who acknowledge the expressive function of punishment reject the idea that this function gives us reason to deprive criminals of their liberty or property.[23] T.M. Scanlon, for instance, writes

> Pointing out "the expressive function of punishment" helps us to understand our reactions to punishing particular kinds of people, but what role if any does it have in the justification of punishment? It seems to have no positive role in justifying hard treatment of the legally blameworthy. Insofar as expression is our aim, we could just as well "say it with flowers" or, perhaps more appropriately, with weeds.[24]

Contra Scanlon, I will argue that when punishment is understood as an expression of indignation, the expressive

[23] In addition to Feinberg, quoted in footnote 12, see, for example, David Boonin, *The Problem of Punishment* (New York: Cambridge University Press, 2008), 176-179; Thom Brooks, *Punishment* (New York: Routledge, 2012), 117-118; Nathan Hanna, "Say What? A Critique of Expressive Retributivism," *Law and Philosophy* 27 (2008): 325-349; H. L. A. Hart, *Law, Liberty and Morality* (New York: Random House, 1963), 66; Victor Tadros, *The Ends of Harm: The Moral Foundations of Criminal Law* (Oxford: Oxford University Press, 2011), 108-109.

[24] T.M. Scanlon, "The Significance of Choice" in *The Tanner Lectures on Human Values, Vol. 8* (Salt Lake City: University of Utah Press, 1988), 214.

function of punishment explains why we have reason to subject criminals to the characteristic deprivations of punishment.

Consider how expressions of another moral emotion, compassion, might be criticized. Imagine a very wealthy man is sitting in front of the television watching coverage of a natural disaster that has been devastating to some members of his community. He tells his personal assistant who is working nearby, "I feel a great deal of compassion for the victims of this disaster. Please send them flowers with a note expressing this compassion." She replies, "If you were really compassionate, you could send them money to help provide for their basic needs and give them an opportunity to rebuild their lives. In fact, if you were really motivated, you could spend some time volunteering at the shelter that has been set up."

There are at least two ways in which the wealthy man's gift of flowers is deficient as an expression of compassion. First, the expression is not, as we might say, "constructive" in light of the values that underlie compassion. When people feel compassion, they judge that the suffering of those for whom they feel it is bad. A purported expression of compassion that does nothing to alleviate this suffering when such alleviation is possible shows a lack of commitment to the value judgment that is characteristic of compassion. Second, a gift of flowers does not express, as we might say, a "proportionate" degree of compassion. Emotions have affective elements that lead to action. Compassion involves a disposition to engage in actions that typically alleviate suffering. Generally, the more intense the compassion is, the more one is willing to help. The wealthy man's refusal to do anything for the victims of the natural disaster beyond sending flowers belies his claim to feel a great deal of compassion. The suffering brought by the natural disaster in this example calls for a high degree of compassion, and merely sending flowers, to put it crudely, does

not express "enough" of it. These mundane observations suggest that in expressing compassion, we cannot always just as well say it with flowers.

Likewise, when it comes to serious crimes, norms of constructiveness and proportionality give us reasons to express our indignation with punishment rather than weeds. In section 1, I suggested that a plausible account of the justification of punishment could not disregard the moral significance of the social goal of crime reduction. The values that make blame important give us reason to express it in ways that contribute to this goal. Just as we characteristically perceive the significance of suffering when feeling compassion, so too does indignation reflect an acknowledgement of the moral consideration that victims of wrongdoing are owed. Where possible, constructive expressions of indignation will reduce the likelihood that others are victimized. Such expressions demonstrate a commitment to the values that partially constitute appropriate indignation.

Punishment's deterrent function makes it a constructive way to express indignation in light of the values that appropriately inspire our indignation. Threatening to deny criminals certain desired goods can deter crime, insofar as the punished and others in society who are aware of the punishment do not want to be denied the good in question. Moreover, the goal of crime reduction partially explains why state punishment takes the form it does. Imprisonment, for instance, incapacitates those who are risks for committing crimes and places them in a situation in which they will have a lot of time to think about their wrongdoing. Prisons often (and should) have programs whereby criminals can earn college degrees and develop skills that will

make it less likely that they will commit future crimes.[25] In these ways, imprisonment increases the likelihood that moral obligations are not violated in the future and, in doing so, protects the rights of members of the community. When we express indignation toward serious crimes through imprisonment, the values that explain the moral importance of that indignation are reflected in the very manner of its expression. This is less obviously true if we express our indignation with weeds. Weeds would not provide much of a disincentive to crime. Weeds would neither incapacitate criminals nor effectively encourage them to reflect on their crimes. We cannot just as well express indignation toward serious crimes with weeds as punishment, because such expressions are not constructive. They do not show the same commitment to the values that animate the reactive sentiments.

Like constructiveness, proportionality gives us reason to express indignation toward the most serious crimes by means of punishment. We understand expressive behaviors that are objectively worse for their targets as signaling those sentiments in a greater degree. In order to signal to all members of the community the value of those who have been victimized, the form that blame takes must be accessible to everyone in society as expressive of an unambiguously high degree of blame. A society that consistently punishes serious crimes emphatically shows that it values the victims of those crimes as being owed moral consideration. Expressions that deprive their targets of something less important could legitimately be viewed by members of society as failing sufficiently to stand up for those victims.

[25] One reasonable worry about the expressive component of punishment is that it stigmatizes the punished. The goal of reducing the likelihood that criminals reoffend gives the state reason to treat them in ways that limit the extent to which punishment permanently ostracizes them from the community.

Resentment and indignation deny their objects something objectively good, a certain sort of social regard. People who feel these emotions are characteristically disposed to perform actions that are bad for their objects, even when those who feel them do not desire that those objects suffer. There are three rough categories in which we might categorize these harms. First are those harms that come with the awareness that someone feels a reactive sentiment toward you. A paradigm example is the distress one feels when one is the target of verbal expressions of blame. Second are the denials of further social goods that come with other people withdrawing goodwill that they would otherwise have toward you. Examples of these harms include being excluded from a social circle or not being able to receive aid with one's projects. Harms in this second category often accompany harms from the first, insofar as those deprivations signal the presence of a reactive sentiment. Third are those harms that are only appropriate when carried out by agents of the state, because they take away goods whose denial requires the threat of force. State punishment denies criminals goods in this category. The third sort of harms often bring the first two with them: when people are imprisoned for committing a serious crime, for instance, they are typically aware that they are being blamed and they are denied the aid of society to some of their life projects.

Holding fixed our relation to the object of our reactive sentiments, the category of harms that our reactive sentiments dispose us to countenance seems to correspond to the degree to which we feel those sentiments. Blaming serious criminals by denying them goods in the third category shows that we endorse a high degree of indignation in response to their crimes. Note that, in practice, our judgments about what makes particular modes of expression proportional indirectly depend on what makes them constructive. Having grown up in societies in which

imprisonment is thought to contribute to crime reduction helps to account for our general disposition to express indignation toward murderers through imprisonment. The deterrent function of punishment thereby informs our view of what a proportionate response is, even in cases in which deterrent aims will not be achieved by an act of punishment.

Given considerations of proportionality and blame's public role as described in the previous section, we have reason to express blame toward those who commit serious crimes by depriving them of goods in the third category. Such expressions are accessible to everyone in society as expressive of a high degree of indignation. We cannot just as well express blame with weeds; weeds would not express a sufficient degree of indignation for this task.[26] Imagine that our government suddenly starts to say it with weeds, sending weeds to convicted murderers rather than punishing them. We and our fellow citizens are likely to have trouble conceptualizing this as genuinely blaming them. Giving someone weeds is not an action that is typically objectively bad for its target and thus is difficult to understand as even a candidate expression of indignation.

But imagine the government makes clear the expressive meaning of these weeds and claims that they are to represent the highest degree of indignation. Those who receive the weeds would presumably be deprived of some benefit of social regard. There may be some in society who would accept this system as expressive of the appropriate amount of blame for murderers, but I suspect most people, given their own emotional experience,

[26] See also Raffaele Rodogno "Shame, Guilt, and Punishment," *Law and Philosophy* 28 (2009): 429-464 at 437 and 459. Rodogno makes a similar point, emphasizing how impositions of the loss of property and liberty are widely believed to embody the emotions of "punitive hostility" in a way that merely conventional symbols of social condemnation would not be.

would find the expression insufficient. Many victims of serious crimes, and those who care about them, will still reflectively endorse their own reactive sentiments that dispose them to approve depriving their perpetrators of goods beyond social regard. A weeds-dispensing system that is not collectively understood to express a sufficient degree of blame will not be capable of discharging its public expressive function. It will seem to many not to stand up adequately for the victims of those crimes, or not to take violations of important moral obligations seriously enough. Giving weeds to murderers would neither convey the degree to which the community is indignant about their actions, nor would it make clear the value that the community attaches to the victim's moral rights and moral standing.

One might worry that this line of reasoning would entail that if members of some community were disposed to express their indignation toward the most serious crimes with cruel and inhumane behavior, norms of proportionality would give the society reason to punish in cruel and inhumane ways, unchecked by legal institutions that protect civil liberties. If this is a defect of the blaming function, however, it is a defect shared by the deterrent and retributive functions. Cruel and inhumane punishment may be an effective deterrent. It is not obvious that cruel ways of inflicting suffering would not carry out the purported retributive function of punishment on criminals who have treated their victims terribly. However, any of these three candidate justifications of punishment could appeal to a wider political theory in response to the objection. As I asserted in the previous section, the blaming function of punishment must be integrated into an account of democratic legitimacy that sets out civil liberties protecting all citizens from certain government actions.

To summarize the past two sections, the reactive theory identifies a positive reason to subject all those who commit serious crimes to proportional punishments. The argument draws on the reactive account of moral responsibility, and in particular on the role of blame and its social expression in protecting and promoting central moral values. The form the expression takes must be capable of communicating a high degree of blame to everyone in society. Punishment is well-suited for this task because it denies the punished goods that we all recognize to be important. Systems of punishment also contribute to crime reduction and thereby demonstrate commitment to the values that make blame appropriate.

IV

The Moral Standing to Object

I have contended that the reactive theory of punishment satisfies the *positive reason* desideratum. In this section, I will maintain that the reactive theory also satisfies the *no valid objection from rights* desideratum: it explains why the punished cannot reasonably object to being denied a good to which they would normally have a right. Using examples of interpersonal blame, where our blaming responses are not mediated by systems of punishment, I will argue that the targets of appropriate blame do not have the standing to object to proportionate blame or its expression, even when that expression involves treatment that would otherwise be morally problematic. The reactive theory interprets state punishment as continuous with interpersonal responses in this respect. Insofar as the deprivation of goods such as liberty and property expresses a proportionate degree of blame for serious crimes, those who are punished for those crimes

cannot object to their punishment on the grounds that it deprives them of liberty or property.

Someone might object to being blamed by those around him because he feels distress when the reactive sentiments are directed toward him. This complaint is reasonable if he was justified in acting the way that he did or has an excuse for acting in that way. It is inappropriate to blame the person in theses cases, because he has not really violated a moral obligation or there is some extenuating circumstance that renders blame inappropriate. In other cases, his objection is unreasonable unless he is properly exempted from blame. Those without capacities for moral reasoning or guiding their behavior in light of moral reasons are appropriately exempted from being the objects of reactive sentiments. But if someone commits an unexcused wrong while in the possession of these capacities, he cannot reasonably object to others, appropriately situated, feeling the reactive sentiments toward him to a proportionate degree. This point is about the standing of those who violate moral obligations to object to blame, and *not* about whether those in a position to blame always have conclusive reason to do so. There are other ways of responding to wrongdoing that are consistent with treating those who commit wrongdoing as responsible agents. For example, the swearing off of the resentment that is characteristic of forgiveness is a way of respecting the forgiven as a moral agent. But people who have committed wrongs are not in a position to demand forgiveness when they are appropriately the object of resentment; as far as resentment is concerned, they are rightly at the mercy of the persons they have wronged.

Not only is it unreasonable for people who have violated a moral obligation to object to others, appropriately situated, feeling the reactive sentiments to a proportionate degree toward them, they also *prima facie* do not have standing to object to the

appropriate expression of those sentiments. Imagine that Jim has betrayed Susie's confidence in a matter of some importance, and she expresses her resentment toward him: "I can't believe you did that! You betrayed my confidence." It would be peculiar for Jim to reply by saying, "You ought not express your resentment like that, because it really hurts my feelings, and hurting feelings is morally wrong." What's inapt about the response, I maintain, is not that the general moral claim is problematic. In general, it is morally wrong to hurt people's feelings, and if Jim had not actually betrayed her confidence, he could legitimately appeal to these feelings in objecting. What makes the response unsuitable is that Jim does not have the standing to object morally to the negative impact an appropriate expression of blame has on him.

One might question this diagnosis. It is unreasonable to object to a wide range of actions performed by others on the basis of hurt feelings. It would be unreasonable, for example, for me to demand that my unrequited love date me, or my teacher give me a good grade, because failure to do so would hurt my feelings. In light of such examples, it might be thought—too quickly, no doubt—that hurt feelings do not constitute much of a basis for objecting to the actions of others. Consider instead, then, expressions of blame in the interpersonal context that take the form of deprivations of goods beyond social regard. Imagine that Susie withdraws from Jim an invitation to a party she is throwing in response to his betrayal. He responds, "You ought not express your resentment like that, because I had a legitimate expectation to go to that party and didn't make other plans that night." This is an unreasonable response, not because it is unreasonable to object to having one's legitimate expectations thwarted, but rather because Jim is not in a position to object to the negative effects that his own wrongdoing has on him when they are a result of his being appropriately blamed for that wrong. In response to Jim's

betrayal, Susie's father might express his indignation by breaking off a mentoring relationship with Jim. Assuming the betrayal was serious enough to make the severing of this relationship appropriate, Jim could not reasonably object that Susie's father has broken a promise to mentor him. Though one can usually reasonably complain about a broken promise, one cannot reasonably demand that promises be kept when doing so would preclude the appropriate expression of blame about something one has done.[27]

Jim's standing to object to similar treatment is not undermined when that treatment is not an appropriate blaming response, however. Were the rescinding of the invitation not a blaming response at all—say a friend of Susie's who is helping throw the party and unaware of Jim's betrayal randomly chooses to disinvite him—Jim could reasonably object to it. Jim also retains the standing to object to blaming responses that are disproportionate. If Susie conspires to get him fired from his job as an expression of her resentment, he could reasonably complain. Such a course of action is not an appropriate expression of blame; it is an excessive response to his wrongdoing, expressing a higher degree of resentment than is called for.

[27] T. M. Scanlon makes a claim that seems to be at odds with my point here. *Moral Dimensions: Permissibility, Meaning, and Blame* (Cambridge: Belknap Press of Harvard University Press, 2008), 142: "Even those who have no regard for the justifiability of their actions toward others retain their basic moral rights—they still have claims on us not to be hurt or killed, to be helped when they are in dire need, and to have us honor promises we have made to them." However, it seems to me uncontroversial that Susie's father is permitted to break his promise to Jim in response to his wrongdoing. This need not be because that promise is somehow conditional on Jim's relationship with Susie. He might have made the promise to mentor before Jim met Susie. He would still be permitted not to honor it in these circumstances.

These observations about the way blame works in the interpersonal context can be extended to explain why serious criminals do not have moral standing to object to punishment that denies them their liberty or property. I argued in the previous section that communities have reason to express indignation toward those who commit serious crime by depriving them of important goods. Such deprivations signal an appropriately high degree of indignation. They show that members of the community take such crimes seriously and are committed to reducing their occurrence. Because punishment is a proportionate way to express indignation toward serious crimes, those who commit such crimes cannot reasonably object to punishment that blames them, even if it deprives them of goods to which they would otherwise have a right. Those who have committed serious crimes and been fairly convicted do not have the same standing to appeal to their rights to liberty and property that other people have, when the deprivation of their liberty and property partly constitute proportionate expressions of blame.[28] It would be unreasonable for a fairly convicted murderer to object at sentencing, "You cannot imprison me; it violates my right to freedom." Murders cannot appeal to their right to liberty, not because this is not a right that people are usually protected by, but because convicted murderers do not have the standing to appeal to that right when members of the community have appropriately endorsed imprisonment as the way to blame those who commit murder. [29]

[28] In order for convictions to be fair, the accused must retain due process rights throughout the criminal justice process, regardless of the rights to which they lose the standing to appeal.

[29] One might want a deeper explanation of what about appropriate blame makes its targets lose standing to object to its negative impact. A candidate explanation might be grounded in the notion that the reactive sentiments respect their objects. Proportionate expressions of blame treat their targets in

Nevertheless, those who commit serious crimes retain their standing to object to deprivations of their liberty and property when those deprivations are not constitutive of appropriate blame. They can validly object, for instance, to random kidnappings on the grounds that they violate their right to liberty. The blaming function of punishment is able to explain why punishment is permitted in just those circumstances in which commonsense morality endorses it: we do not think serious criminals forfeit their rights without condition, but only when they are targets of proportionate punishment that blames them for their crimes.

V

Conclusion

The secure place in our moral thinking of the conviction that those who commit serious crimes should be punished is accounted for by the reactive theory of punishment. First, the theory satisfies the *positive reason* desideratum: we think that there is positive reason to punish such wrongs, because blame is called for in response to serious wrongs, and punishment is an

ways that would usually fail to respect them in ways that all persons are owed, but blaming expressions always respect their targets as moral agents. Mitchell Berman recommends a strategy like this one to the retributivist in "*Punishment and Justification,*" *Ethics* 118 (2008): 258-290 at 279. This suggested deeper explanation behind the moral standing account of the permissibility of punishment relies on a questionable premise, however: that the only valid grounds that someone could have to object to a form of treatment is that the treatment fails to respect his agency. The reasons why expressions of blame are inappropriate are hardly exhausted by such failures. Some excessive expressions of blame address their targets as responsible agents, for instance. I believe that the notion of the moral standing to object to blame is best understood as fundamental and not analyzable in other terms.

appropriate way to express this blame. Second, the theory satisfies the *no valid objection from rights* desideratum: we do not think that someone who has committed a serious wrong has the standing to reasonably object to an appropriate expression of blame on the grounds that it deprives him of a good to which he has a right.

In addition to satisfying these desiderata, the reactive theory embodies some of the features of our moral thinking that make deterrence theory and retributivism plausible in the first place. I will conclude by suggesting how the reactive theory is able to take on board these features while satisfying the *positive reason* and *no valid objection from rights* desiderata. Reduction of instances of serious wrongdoing is clearly a morally important social goal. The deterrent function of punishment is part of the reason that communities decide to spend their limited resources on justice systems and detention facilities. Theories of punishment that ignore this function overlook something of moral importance. The reactive theory is not guilty of this sin of omission. According to the reactive theory, part of the reason why blame takes the form of punishment is that punishment deters crime. Because deterrence promotes those values that underlie the community's indignation, punishment that deters crime is an especially constructive expression of the community's indignation. Nevertheless, on the reactive theory, a particular instance of punishment need not actually deter in order to be justified. A successful expression of blame is all that is necessary. Even if imprisoning the assassin of the civil rights leader described in section 1 would serve no deterrent function, for instance, it would proportionately express indignation in our actual social context.

The reactive theory of punishment also satisfies the *no valid objection from rights* desideratum. According to the reactive theory, justified punishment just *is* appropriate blame, and the targets of

appropriate blame do not have the moral standing to object to such blame. This understanding constrains the deterrent goals of punishment. Punishment is justified, on the reactive theory, only if, and to the extent that, it expresses an appropriate degree of indignation in response to a serious crime. We cannot punish people merely as a means to social order. Even if a given punishment would reduce crime, the punished can reasonably object to it if deprives them of a greater good than would express a proportionate degree of indignation.[30]

The other traditional justification of punishment, retributivism, does not justify punishment based on its consequences, but instead based on the idea that it is a fitting response to wrongdoing. The notion that punishment is a backward-looking response to wrongdoing captures something central to our conception of it, acknowledging the agency of the wrongdoer. The reactive theory also justifies punishment as a response to wrongdoing but departs from retributivism on the response that is called for. Retributivism holds that wrongdoers deserve to suffer in proportion to the harm they have caused their victims, and that punishment inflicts this deserved suffering. The reactive theory holds that serious wrongdoers should be held accountable for their wrongs, and that punishment is an appropriate way to blame the most serious crimes. This difference between the responses called for on the two theories is key to explaining why the reactive theory avoids the difficulties of retributivism in satisfying the *positive reason* and *no objection from rights* desiderata.

[30] Likewise, while norms of constructiveness give us reason to express blame in ways that will lead the punished to reform their behavior, the expressive function of punishment is carried out in cases where the offender is insensitive to the moral opprobrium of the community. The indignation of society is still expressed in a way that shows that it attaches value to the offender's victim.

In section 1 I argued that reasonable skepticism about the retributive idea undermines retributivism's claim to satisfy each of the desiderata. The reactive theory is based on the "reactive idea" that the reactive sentiments are appropriate responses to wrongdoing. Rather than being "a mysterious piece of moral alchemy," the fittingness of these sentiments is a lived conviction for most of us, given to us with our involvement in interpersonal relationships. I have offered an account of why these sentiments survive moral reflection: we endorse the value judgments that partially constitute them concerning the moral status of the person who has been wronged. The reactive sentiments are not vindictive and do not characteristically involve the thought that it would be intrinsically good for their object to suffer. Justified reactive punishment does not aim to inflict suffering on the punished for its own sake. Instead, it gives appropriate expression to the justified indignation of the community, in a way that is constructive in light of the values that constitute the community's indignation.

I raised a further concern about whether retributivism satisfies the *positive reason* desideratum: it does not seem to explain we have positive reason to punish serious criminals in cases where their punishment does not cause them to suffer more than they would have suffered were they not punished. If, on the other hand, we punish in order to express blame, there is nothing puzzling about these cases. The reactive sentiments do not necessarily involve the idea that their object should suffer, so blame can be expressed effectively in ways that do not in fact turn out to inflict suffering. According to the reactive theory, we have reason to express the community's indignation toward serious crime through punishment because it publicly shows that we take certain violations of moral obligations seriously and take the victims of crime to be owed moral consideration. This expression can be

successful regardless of whether it causes its target to suffer. Sending criminals to prison expresses a high degree of indignation whether or not they suffer a great deal day to day as a result. The reactive theory can explain why we have positive reason to punish even in those cases in which punishment will not actually inflict suffering on the punished.[31]

Illinois Wesleyan University

[31] Thanks to audiences at the University of California-Berkeley, Illinois Wesleyan University, The Chinese University Hong Kong, and the Philosophy Club at the University of Illinois at Urbana-Champaign. Special thanks to Markus Kohl, Niko Kolodny, Samuel Scheffler, R. Jay Wallace, and two anonymous reviewers at *Philosophy and Public Issues*.

If you need to cite this article, please use the following format:

Engen, Andrew, "The Reactive Sentiments and the Justification of Punishment," *Philosophy and Public Issues (New Series)*, Vol. 5, No. 1 (2015), 173-205, edited by S. Maffettone, G. Pellegrino and M. Bocchiola