1-1-2015

A New Reading of Kant's Theory of Punishment

Robert Hoffman
University of Pennsylvania, robert.hoffman.22@gmail.com

A New Reading of Kant's Theory of Punishment

Abstract
There are deep, insurmountable difficulties with the traditional interpretation of Immanuel Kantâ??s writings on the subject of punishment. Although it is undeniable that throughout his published writings on practical philosophy â?? and in particular in his Metaphysics of Morals â?? he consistently advocates for the view that punishment can only be justified as a direct response to an individualâ??s act of wrongdoing, his status as one of the foremost theorists in the retributivist pantheon is philosophically untenable. In this dissertation, I articulate the ways in which Kantâ??s explicit support for retributivism directly contradicts more foundational elements of his practical philosophy and argue instead that he has the resources to consistently construct a deterrent theory of punishment. In particular, I highlight Kantâ??s division of duties and his conception of the state to demonstrate that the idea of a political community retributively responding to moral desert is wholly incompatible with Kantian principles. In order to overcome these obstacles, I develop a new approach to Kantian deterrence â?? which I call Kantian Protective Deterrence â?? that grounds the stateâ??s right to exercise coercive force against its citizens in what Kant understood to be its fundamental role of protecting each individual citizen from violations of her or his right to exercise external freedoms.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Philosophy

First Advisor
Paul Guyer

Keywords
Immanuel Kant, Punishment

Subject Categories
Philosophy

A NEW READING OF KANT'S THEORY OF PUNISHMENT

Robert Hoffman

A DISSERTATION

in

Philosophy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

_____

Paul Guyer
Jonathan Nelson Professor of Humanities and Philosophy

Graduate Group Chairperson

_____

Michael Weisberg
Associate Professor of Philosophy

Dissertation Committee
Samuel Freeman, Avalon Professor in the Humanities, Philosophy
Adrienne Martin, Akshata Murty '02 and Rishi Sunak Associate Professor of Philosophy
Charles Larmore, W. Duncan MacMillan Family Professor in the Humanities
Karen Detlefsen, Associate Professor of Philosophy and Education

A NEW READING OF KANT'S THEORY OF PUNISHMENT

For my parents.

I am indebted to a great many people for their assistance, advice, encouragement and support over the course of this project. A dissertation is a learning process in more ways than one, and I have been exceptionally fortunate to have so many remarkable teachers. Paul Guyer – my supervisor – not only guided the development of the projeect, but also my development as a philosopher. I also owe a great debt to Samuel Freeman and Adrienne Martin, who served on my committee since its beginning, and to Charles Larmore, who generously contributed to the project the last few years. Their insight and feedback helped to shape the direction and scope of the dissertation, and it is immeasurably better for their input. Finally, I must thank Karen Detlefsen, who regularly went well beyond the call of duty when I was most in need of assistance or guidance.

It is hard to overstate the amount I have learned from and relied on my fellow graduate students over the past six years. Chris Melenovksy and Justin Bernstein helped sharpen and refine my understanding of political philosophy, and Wiebke Deimling's insight into Kant has likewise proved invaluable. I am also forever indebted to Emily Parke, Rob Willison, and Hal Parker, whose friendship and support from my very first day of graduate school made this process far more enjoyable and rewarding than it would have otherwise been.

Finally, I would be remiss if I did not recognize that producing a dissertation is as much a personal undertaking as it is an academic one. As such, I am indebted to my parents and to my brothers, John and Michael. And, of course, I cannot possibly describe all I owe to Heather, who not only made any of this possible, but also worthwhile.

ABSTRACT

A NEW READING OF KANT'S THEORY OF PUNISHMENT

Robert R. Hoffman

Paul Guyer

There are deep, insurmountable difficulties with the traditional interpretation of Immanuel Kant's writings on the subject of punishment. Although it is undeniable that throughout his published writings on practical philosophy – and in particular in his *Metaphysics of Morals* – he consistently advocates for the view that punishment can only be justified as a direct response to an individual's act of wrongdoing, his status as one of the foremost theorists in the retributivist pantheon is philosophically untenable. In this dissertation, I articulate the ways in which Kant's explicit support for retributivism directly contradicts more foundational elements of his practical philosophy and argue instead that he has the resources to consistently construct a deterrent theory of punishment. In particular, I highlight Kant's division of duties and his conception of the state to demonstrate that the idea of a political community retributively responding to moral desert is wholly incompatible with Kantian principles. In order to overcome these obstacles, I develop a new approach to Kantian deterrence – which I call Kantian Protective Deterrence – that grounds the state's right to exercise coercive force against its citizens in what Kant understood to be its fundamental role of protecting each individual citizen from violations of her or his right to exercise external freedoms.

Table of Contents

1 Kantian Protective Deterrence: An Introduction

Immanuel Kant is frequently hailed as the foremost philosopher of the Enlightenment. His work in moral philosophy has inspired a tradition of thinking that still stands as one of the dominant approaches to answering normative questions. The values of unconditional respect for human persons, strict commitment to inviolable rights, and the intrinsic worth of autonomy and agency – all hallmarks of the fragile but precious moral progress in modern times – find unequivocal support and foundation in his writings. The influence of Kant's insight, originality, and breadth of thought is almost impossible to overstate.

His prominence and significance within the history of moral philosophy alone would warrant a careful consideration of Kant's views on the subject of punishment. As it happens, however, there is additional reason to study his position: namely, Kant's reputation as one of the central, foundational philosophers of the retributivist school of thought. Retributivism – the belief that punishment is justified as a response to the wrong actions and moral desert of the perpetrator – has persisted as one of the dominant theories of punishment, from the earliest points in human history to today. Kant, in turn, is regularly singled out as the first philosopher in modern times to provide a secular, rather than religious, basis for retributivism. So great is his identification with the retributive view that it is sometimes merely labeled the deontological approach to punishment.

Kant's reputation as a theorist of retributivism is well-earned. In his *Metaphysics of Morals*, he describes punishment as being justified only when it is an immediate response to an act of criminal wrongdoing: "Punishment can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society. It must always be inflicted upon him only *because he has committed a crime*" (6:331).[1] The state's response to criminal wrongdoing is analytically necessitated (27:552) and should take the form of *ius talionis* (the law of retribution): an eye for an eye (6:332). Failure to respond in this way makes a state complicit in the act of wrongdoing, even in such an extreme condition as the state's dissolution (6:333). These views are expressed throughout the *Metaphysics of Morals*, but also receive attention in his essay "On the Common Saying: That May Be Right in Theory but Not in Practice," as well as in his lectures on moral and political philosophy.

There is ample evidence to demonstrate Kant's commitment to retributivism. When it comes to how he defends his position, however, there is considerably less to be said. While Kant offers some support for his view, this support is insufficient for a variety of reasons, the most fundamental of which is a failure to ever demonstrate why the state is authorized to respond to the purported moral desert of its citizens. Kant's retributivism ultimately comes unmoored from the rest of his practical views and threatens to cast them all into deep inconsistency. While I believe any such effort is

[1] Kant, Immanuel. *Practical Philosophy*. Ed. and trans. Mary J. Gregor. Cambridge: Cambridge University Press, 1996, and *Lectures on Ethics*. Ed. J.B. Schneewind and Peter Heath, trans. Peter Heath. Cambridge: Cambridge University Press, 1997. All internal citations refer to the standard Prussian Academy edition, volume and pages.

likely to fail, one might respond to this looming contradiction by attempting to construct a new Kantian justification for retributivism. Instead, this project will show that a mixed theory of punishment with a deterrent justification is not only available to Kant, but it is more consistent with his underlying practical philosophy than any retributive theory could hope to be.[2]

In addition to uncovering the nuances of Kant's views and critically assessing their consistency, this dissertation will also endeavor to say something about the nature of punishing itself. Put another way, this is a project both about historical interpretation and about our practices of punishing today. We live in a time and place characterized by hard questions concerning both the morality and the efficacy of our institutions and practices of punishing. While Kant is not the only theoretician in the retributivist camp, he provides one of its foremost philosophical foundations. By interrogating what it would take for Kant truly to be a retributivist – and what it would mean for retributivism if Kant cannot truly sustain the position – I hope to make a meaningful

[2] Traditionally, the standard triumvirate of justifications offered for punishment consisted of retributivism, deterrence, and rehabilitation. Retributivism, characterized as a backward-looking justification, describes punishment as a response to the moral desert of an agent who had previously committed a wrong. Deterrence, characterized as a forward-looking justification, rose to prominence in the early modern period and led to substantial reforms to criminal justice systems across Europe and North America through the nineteenth century. According to deterrence, punishment is justified as a way of discouraging future acts of crime and thus promoting the general welfare. Rehabilitation, which justifies punishment as a means of correcting the deficiencies in the character of the person who committed the crime, was most prevalent among ancient Greek philosophers, and while many theories describe rehabilitation as an admirable goal, few defend it as a sufficient justification in its own right. In recent years, other efforts to justify punishment have been made, including appeals restitution, self-defense, and security. As Kant was only familiar with the traditional trinity, however, I will be confining the focus of this dissertation to retributivism, deterrence, and rehabilitation.

contribution to the contemporary conversation about why and how we should and should not punish.

1.1 A Changing Scholarship

The topic of Kant's theory of punishment has received a growing and more critical quantity of attention in recent years. As Anglophone scholarship begins to take more seriously Kant's later works on political philosophy – including primarily the *Metaphysics of Morals*, as well as "On the Common Saying" and "Toward Perpetual Peace" – the question of how Kant explains the state's authorization to punish its citizens becomes an increasingly compelling and important one. Efforts to address this question have taken the form of numerous articles and chapters in books; to date, there are no monographs dedicated to the question of Kant's theory of punishment. As such, there is some need for a more comprehensive analysis of the subject than is possible in the length afforded by an article or chapter.

Within the recent scholarship, there are three dominant trends that can be identified. The first and most prevalent trend is a move toward challenging the traditional narrative of Kant as the grandfather of retributivism. Instead, these works argue, there are strong themes and undercurrents of deterrence running throughout Kant's political philosophy. I will be describing works of this sort as defending a 'Kantian deterrence' view, for obvious reasons. Although Kantian deterrence enjoys an increasingly dominant place in Kant scholarship, there are still numerous varieties of deterrence and interpretative divisions within this group. The second major trend within

the recent literature on Kant's theory of punishment is a move to defend the orthodox reading of Kant. Kant interpreters who are producing work with this aim and focus – what I will be calling 'orthodox retributivism' – are motivated by an interest in showing that Kant's retributivism can respond to the challenges raised by the deterrence theorists. The third and final broad trend that can be identified in the recent literature is a small but important rejection of the possibility of a consistent Kantian position on the question of punishment. I will be referring to arguments and positions of this sort as the 'anti-theory' view. According to the anti-theory view, all efforts to find in Kant or construct on his behalf a coherent theory of punishment are doomed to end in contradiction and failure.

The Kantian deterrence theorists are motivated by two primary considerations. First, Kant interpreters have recently begun exploring the ways in which deterrent elements are, in fact, built into Kant's writings. Papers of this sort posit that strict interpretation of Kant can uncover enough evidence to conclude that Kant's theory of punishment is, at least partially, deterrent. The foremost of those defending positions of this kind is Sharon Byrd, whose paper "Kant's Theory of Punishment: Deterrence in Its Threat, Retribution in Its Execution"[3] is heralded by virtually all those working on Kant and punishment as the seminal reexamination of Kant's views on punishment. Byrd's model for incorporating both deterrence and retribution in a Kantian theory has been a strong inspiration for many, and any work on Kant's theory of punishment will need to

[3] Byrd, B. Sharon. "Kant's Theory of Punishment: Deterrence in Its Threat, Retribution in Its Execution." *Law and Philosophy*, Vol. 8, No. 2 (Aug., 1989), pp. 151-200

grapple with her paper (as mine does in chapter four). Other important interpreters who explore Kantian deterrence views in this vein include Paul Guyer,[4] Allen Wood,[5] Arthur Ripstein,[6] Nelson Potter,[7] Sarah Holtman,[8] and, in some cases, Thomas Hill.[9]

The second major consideration motivating Kantian deterrence is a general sense among Kantians that the nature of retributivism is deeply at odds with other, more fundamental elements of Kant's practical philosophy. This motivation can be seen in the work of Thomas Hill,[10] Don Scheid,[11] Arthur Shuster,[12] Carol Steiker,[13] and Matthew Altman.[14] Unlike the group described above, these deterrence theorists take Kant to be a committed retributivist. Nevertheless, they argue that for reasons of consistency, he ought not to have embraced retributivism. Unlike the previous group, then, these Kantian deterrence theorists view deterrence as something that must be imputed to Kant, rather than uncovered within his writing. In spite of this, however, they consider

---

[4] Guyer, Paul. *Kant*. 2nd ed. London: Routledge, 2014.
[5] Wood, Allen W. *Kantian Ethics*. Cambridge: Cambridge University Press, 2008.
[6] Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge: Harvard University Press, 2009.
[7] Potter, Nelson. "Kant on Punishment." *The Blackwell Guide to Kant's Ethics*. Ed. Thomas E. Hill, Jr. London: Blackwell Publishing, 2009
[8] Holtman, Sarah. "Toward Social Reform: Kant's Penal Theory Reinterpreted," *Utilitas*, 9 (1997), pp. 3-21
[9]Hill, Thomas E., Jr. "Kant on Wrongdoing, Desert, and Punishment." *Law and Philosophy*, Vol. 18, No. 4 (Jul., 1999); and "Treating Criminals as Ends in Themselves." *Jahrbuch fuer Recht und Ethik*, Vol. 11 (2003), pp. 30-31.
[10] Hill, Thomas E., Jr. "Kant on Wrongdoing, Desert, and Punishment." *Law and Philosophy*, Vol. 18, No. 4 (Jul., 1999), pp. 407-441; "Kant on Punishment: A Coherent Mix of Deterrence and Retribution?" *Jahrbuch fuer Recht und Ethik: Annual Review of Law and Ethics* 5, (1997), pp. 291-314
[11] Scheid, Don. E. "Kant's Retributivism." *Ethics*, 93 (1983), pp. 262-282.
[12] Shuster, Arthur. "Kant on the Role of the Retributive Outlook in Moral and Political Life." *The Review of Politics*, Vol. 73, No. 3 (SUMMER 2011), pp. 425-448
[13] Steiker, Carol S. "No, Capital Punishment is Not Morally Required: Deterrence, Deontology, and the Death Penalty."
[14] Altman, Matthew C. "Subjecting Ourselves to Capital Punishment: A Rejoinder to Kantian Retributivism." *Public Affairs Quarterly*, Vol. 19, No. 4 (Oct., 2005), pp. 247-264.

deterrence to be compatible with Kant's other moral and political positions. Although there is some overlap in their theories – for instance, in the work of Hill and Scheid – the main difference within the work of these interpreters is the form and nature of deterrence that they see as being ideally suited to the other aspects of Kant's practical philosophy. The version of deterrence that enjoys the widest support in this group is one that grounds the state's authorization to punish in the "hindering to a hindrance to freedom" argument. It is within this broad interpretative category that my position will ultimately rest.

Opposed to the Kantian deterrence views that have emerged recently are a smaller but no less forceful group of papers defending the orthodox retributivist reading of Kant. While some of these are efforts to explain the centrality and necessity of retributivism for Kant's practical philosophy, [15] most others are direct responses to the deterrent challenges raised by Byrd, Hill, Ripstein, Scheid, and others. The foremost of the scholars working in this trend is Jeffrie Murphy. [16] Working within both history and philosophy of law, Murphy is an ardent defender of retributivism in general and Kant's version in particular. Although Murphy has recently begun to express doubts about the feasibility of Kantian retributivism (discussed below), his earlier work still represents one of the strongest efforts to defend the traditionalist position.

---

[15] Parrish, John M. and Tuckness, Alex S. "Kant and the Problem with Pardons." *Western Political Science Association*, Annual Meeting. (March 31, 2010).

[16] Murphy, Jeffrie G. *Kant: The Philosophy of Right*. Macon: Mercer University Press, 1970.; "Kant's Theory of Criminal Punishment." *Retribution, Justice, and Therapy: Essays in the Philosophy of Law* (Dordrecht, Holland: D. Reidel, 1979), pp. 82-92; "Three Mistakes about Retributivism." *Analysis*. Vol. 31, No. 5 (Apr., 1971).

Samuel Fleischacker responds to the challenges of the deterrence theorists in a different manner.[17] Fleischacker articulates a reading of Kant that is still – in my view – retributivist, but it is a considerably more articulate, complex version of retributivism than the orthodox understanding of Kant. Unlike Murphy – who attempts to defend retributivism on legal and political grounds – Fleischacker bases his defense in Kant's moral philosophy. In short, Fleischacker's approach is to focus on the role of maxims and the Formula of Universal Law in grounding punishment retributively. In the fourth chapter of the dissertation, I will argue that Fleischacker's approach still cannot solve Kant's difficulties with the 'hard problem' of retributivism: namely, why the state is authorized to respond coercively to some – but not all – instances of the citizens' moral desert.

The final major scholarly trend to consider is the denial that Kant has the ability to articulate and defend a coherent theory of punishment. I have described this as the anti-theory view. Support for the anti-theory stems largely from the same sources as Kantian deterrence, but for one crucial difference: it denies the claim that Kant has the resources to ground a deterrence theory. Instead, Kant's retributivism fails as a coherent theory of punishment, and nothing is left to replace it.

Perhaps the most noteworthy example of an anti-theory pessimist is Murphy.[18] Although Murphy was previously one of the staunchest contemporary defenders of the

---

[17] Fleischacker, Samuel. "Kant's Theory of Punishment." *Essays on Kant's Political Philosophy*. Chicago: University of Chicago Press, 1992, pp. 191-212.

[18] Murphy, Jeffrie G. "Does Kant Have a Theory of Punishment?" *Columbia Law Review*, Vol. 87, No. 3 (Apr., 1987), pp. 509-53

retributivist Kantian orthodoxy, he cites the work of Byrd and others as raising decisive objections to such a position. He does not accept the deterrence position posited by the Kantian deterrence theorists however; instead, he argues that any such position is ruled out by the constraints imposed by the Formula of Humanity. This leads him to conclude that there is no 'Kantian' theory of punishment.

A second variant of theory pessimism can be found in a paper by Jean-Christophe Merle.[19] Although Merle begins his paper with a defense of sorts of Byrd and the deterrence readings of Kant, he ultimately expresses reservations about such readings and distances himself from them. While he does construct a theory with elements inspired by Kant's practical philosophy, Merle does not think that Kant himself has the means to put together a consistent account of punishment.

The clashes between Kantian deterrence, orthodox retributivism, and anti-theory pessimism provide a rich and complex background for this dissertation. In working out a novel, distinct position, I will address many of the most prominent and influential views above. All three of the major trends will appear throughout the project, but the various versions of Kantian deterrence will play the most significant role. The challenge for my project will be to demonstrate 1) that Kantian practical philosophy can sustain a robust, consistent theory of punishment, 2) that the best version of such a theory will incorporate deterrent, retributive, and rehabilitative elements, but will ultimately rely on a deterrent justification, rather than a retributive one, and 3) that I have an original

[19] Merle, Jean-Christophe. "A Kantian Critique of Kant's Theory of Punishment." *Law and Philosophy*, Vol. 19, No. 3 (May, 2000), pp. 311-338

approach to Kantian deterrence that avoids some of the potential difficulties facing other, existent deterrent theories.

1.2 A New Direction

Broadly stated, I contend that the Kantian deterrence movement is correct: a theory of punishment with a deterrent justification is more consistent with the fundamental aspects of Kant's practical philosophy. Although Kant himself was clearly a committed retributivist who rejected the moral permissibility of deterrence as a justification for punishment, he asserted this retributivism on unstable and indefensible foundations and overstated the dangers and difficulties of deterrence. His reliance on a relatively conventional liberal political philosophy commits him to a conception of the state and its purpose that leaves open the possibility of adopting a deterrent justification. Indeed, such a justification can be made compatible with his moral philosophy with relatively minor adjustments.

I do not intend, however, to merely defend the versions of Kantian deterrence that have been developed up to this point. Instead, I will articulate and defend a position that I call '**Kantian Protective Deterrence**.' According to Kantian protective deterrence, the state's purpose is to make determinate and preserve for its citizens an equal, maximally comprehensive scheme of rights and external freedoms. This purpose underlies the state's permission to adopt certain measures, constrained by moral and political principles, to reduce and prevent any violations of the citizens' rights and exercise of their external freedom. The threat and subsequent execution of punishment is

one of these measures. Unlike many of the dominant varieties of deterrence, which justify state punishment on the grounds that crime represents a threat to the civil order, the continued existence of the state, or the supremacy of the state's authority, Kantian protective deterrence justifies deterrent measures – such as punishment – simply by reference to the state's obligation to protect each individual citizen from violations of her or his rights. In this way, my position aligns itself less with distributive justice, and more with the tradition of commutative justice.[20] Rather than justifying punishment by reference to some advantageous social arrangement or the intrinsic, non-instrumental value of the state, I will do so by reference to the rights of individuals.

Unlike some other deterrence views,[21] Kantian protective deterrence is not primarily concerned with attempting to show that those who engage in criminal wrongdoing are not morally deserving of suffering. Although I think these arguments have some merit, Kantian protective deterrence is prepared to grant that wrongdoing might be analytically connected to moral desert. Even if this is the case, however, I contend that Kant is best served by a deterrent theory, in light of his failure and inability to provide any strong reason for why the state is authorized or required to respond to such moral desert. In making this argument, I will be responding primarily to the work of the orthodox retributivists.

---

[20] The concept of commutative justice can be traced to Aristotle's *Nicomachean Ethics*. As opposed to distributive justice, which concerns the arrangement of goods and resources within a society or state, commutative justice is focused solely on the interactions and rights of individual citizens.

[21] See, for example, Hill (1997) and Wood (2008).

Another key feature of Kantian protective deterrence is its incorporation of both retributive and rehabilitative elements within a broadly deterrent framework. This approach marks it as what is called a 'mixed' theory of punishment in contemporary literature on criminal justice. I will say more about the nature of mixed theories and the precise way in which deterrence, retribution, and rehabilitation are combined below, but for now suffice it to say that while the theory is committed to deterrence as the sole justification for punishment, the application and functioning of the institutions of punishment are constrained by retributive and rehabilitative interests.

Kantian protective deterrence also has the advantage of developing a full, comprehensive analysis of every aspect of punishment. The greater scope afforded by a project of this length allows me to explore not only questions of the justification of punishment and significance of *ius talionis*, but also a set of lesser explored questions surrounding Kant's theory of punishment, including the role played by the division of duties Kant establishes in the *Metaphysics of Morals*, the permissibility of rewards, the methods of punishment that are and are not acceptable, the possibility of rehabilitation, the subject of international courts and punishment, and the possibility of morally permissible but legally punishable acts of civil disobedience. By exploring this comprehensive range of questions, I aim to show that not only is Kantian protective deterrence consistent with his underlying practical philosophy, but it can also do important work in answering practical questions in detail.

Throughout the dissertation, I employ an interpretative methodology that is broadly reconstructive and governed by two ordered principles. In describing my

approach as reconstructive, I mean to both convey a positive picture of the methodology of the project, as well as distinguish it from other alternative methodologies employed by those working within the history of philosophy. The dissertation aims at building a coherent, internally consistent theory out of Kant's philosophy as it is expressed in his published works and the notes on his lectures preserved by his students. This effort, however, is not committed exclusively to uncovering the most faithful representation of all the details of Kant's views. Instead, it seeks to select the most successful of these details and craft them into a unified view. In doing so, I hope to occupy a middle position between those engaging in strict interpretation and the kind of 'Kantian' view that takes its inspiration from some small set of Kantian principles but develops them independently of any historical concern for Kant's own views.

This reconstructive project is guided by two interpretive principles: 1) examine and endorse Kant's most foundational philosophical commitments; and 2) attempt to retain as many of his explicit statements about punishment as possible, where this does not violate the first principle. Although not all of Kant's thoughts on punishment can be preserved by such an interpretive strategy, the ones that are excluded are ruled out on the basis of their inconsistency with Kant's more basic writings on moral or political philosophy. The end result of this interpretation is a theory that, despite being different from Kant's own, is still Kantian in the sense that it is constructed from within the

framework of his fundamental commitments and still endeavors to preserve as much of his original view as possible.[22]

1.3 A 'Theory' of Punishment

In addition to the above interpretative methodology, this project is also guided by a specific theoretical framework that contributes to both the structure and content of the dissertation. Specifically, I analyze Kant's writings on punishment under a very precise conception of what a 'theory of punishment' is. It is easy to focus so closely on the concept of punishment that one can lose sight of what it means for an account to be a *theory* of punishment at all.

For the purposes of this dissertation, I will be taking a theory of punishment to be a philosophical account comprised of a determinate number of discreet elements. Any fully realized theory of punishment must include five components: a definition, a justification, principles of liability, specification of amount, and criteria for selecting the methods of punishment. Each component is an answer to a different question.[23] It is my

[22] The rationale behind this approach relies on a particular motivation for investigating the history of philosophy. While I am interested in investigating the ways that historical positions and arguments bear on current issues in philosophy, I hold that any value that these investigations will have for illuminating contemporary questions is entirely contingent upon a detailed and accurate grasp of the historical views in question. As such, my goal is to develop a Kantian theory that is robust enough to speak to contemporary concerns about the practice and institutions of punishment, while at the same time preserving the core of what is distinctive in Kant's practical philosophy. I contend that the interpretive strategy I have outlined is the best way to achieve these goals.

[23] E.g., liability answers the question "Who should be punished?" while amount answers "How much punishment is appropriate?"

contention that Kant offers answers to each of these questions, and that Kantian protective deterrence can do the same.

There are many ways in which a position on or account of punishment can fail to be a proper theory. The absence of some elements would render a theory incomplete, while the absence of others would make it impossible to describe a position as a theory at all. Likewise, even a theory that contains all the necessary elements can still fail by arranging these elements in a contradictory, unsustainable way.[24] While not all elements of a theory must aim at the same goods, concerns, or interests, they must be structured in an ordered, harmonious way.

These necessary elements of a theory of punishment are deeply inspired by HLA Hart's division of punishment, outlined in his collection of essays, *Punishment and Responsibility*.[25] Here, Hart defends the possibility of 'mixed' theories of punishment by separating a theory's 'general justifying aim' from its 'principles of distribution.'[26] According to Hart, any theory must offer a definition of punishment, must explain the aim that justifies the practice or institution of punishment, and must provide principles of distribution. This last category includes the liability and amount of punishment. In drawing distinctions in this way, he endeavors to establish the possibility of a theory justifying punishment according to one kind of aim, while specifying principles of distribution in accordance with some other.[27]

---

[24] It is this particular failing that most of the 'anti-theory' advocates accuse Kant of.
[25] Hart, H. L. A. *Punishment and Responsibility: Essays in the Philosophy of Law.* 2nd ed. Oxford: Oxford University Press, 1970.
[26] Ibid., p. 4.
[27] Ibid., pp. 8-10.

Although Hart outlines this conception of a theory of punishment almost two centuries after Kant's writings on the subject, there is still room for Hart's schematization to influence a project on Kantian punishment. Although it would be anachronistic to expect Kant to have conceived of a theory of punishment according to the kinds of divisions Hart defends, it is still possible for us to use them as a map of the conceptual space. Kant has something to say about each of the various elements, and although he does not always see them as separable in the way that Hart does, viewing them in this manner allows for us to understand the ways in which the various aspects of the theory interrelate.

In developing my own conception of the constitutive elements of a theory of punishment, I have made several modifications to Hart's schema. In particular, I have added one additional component and divided another into two discreet parts. A full theory of punishment, then, is comprised of five distinct elements: a definition, a justification, principles of liability, specification of amount, and criteria for selecting the methods of punishment.

First, a theory of punishment must provide a **definition** of punishment. The definition spells out what the necessary and sufficient conditions are for an act of violence or coercion to be punishment. There are many kinds of justifiable violence, but not all of them can be understood as punishment; self-defense, for instance, is clearly not punitive. Harder cases can include the "punishment" of children by parents or the actions of a state during a civil war. The very same action might be justifiable as punishment but not justifiable as some other act, or vice versa. We need a definition of

punishment in order to know which kinds of actions can be grouped into this category. According to Kantian protective deterrence, Kant's definition of punishment is a coercive action, undertaken against a citizen of a state by the legitimate executive, as a sanction for the violation of public law.

Second, a theory of punishment must offer a **justification** for punishment. The justification explains why the class of actions picked out by the definition is morally or politically permissible. This justification may make reference to some underlying moral or political principles, but insofar as we think that the category of punishment picks out some importantly distinct set of actions, we should consider the possibility that it is justified in a non-reducible way. In many respects, the justification is the most important element of a theory of punishment, as it explains the reason for all of the rest of the features of a theory. It is the justification that gives the organizing force to the rest of the theory. For instance, theories that justify punishment by reference to retribution, desert, and rehabilitation would be described as a retributive theory, deterrence theory, and rehabilitative theory, respectively, regardless of how the other elements of the theory are specified. It is for this reason that Kantian protective deterrence, despite including elements of retributivism and rehabilitation, can still be appropriately described as a deterrence-based theory.

Third, a theory of punishment needs to specify the **liability** of punishment. In other words, it needs to indicate who is an appropriate target of punishment. While one might think that this question is also answered by the justification a theory adopts, there

might well be other kinds of constraints that must be taken into account.[28] A theory's liability, then, is the specification of the appropriate targets of punitive actions, in light of all relevant details and constraints. I will argue that Kant's original use of retributivism to fix the liability of punishment ought to be preserved within Kantian protective deterrence.

Fourth, any theory of punishment must specify the **methods** of punishment that are appropriate. The specification of the appropriate methods takes two forms. First, the theory must have some general class of acceptable methods of punishing. In establishing such a category, the theory would also necessarily rule out certain methods. We might imagine, for instance, a theory that identifies imprisonment, fines, and mandatory community service as acceptable methods of punishing, but rules out torture and execution as acceptable methods. Second, in specifying the methods of punishment, a theory must also provide some means of determining which method is fitting in particular instances of crime. It is possible that multiple methods are an acceptable response to a particular crime; we might think that either imprisonment or steep fines are fine ways of punishing an act of assault. As long as the theory can give some rationale for which methods are acceptable in a particular instance and which are not, it

[28] For instance, take a theory with a retributive justification of punishment. Under such a theory, punitive actions are justified just in those cases where the target has committed a punishable action. In this case, we might think that a theory with a retributive justification would necessarily specify liability in a way that identifies all and only those who have committed punishable actions as appropriate targets of punishment. This, however, is not the whole story; there may be mitigating factors (such as mental health or extreme circumstances) that would exculpate one who is otherwise an appropriate target of punishment. Liability, then, must take into account any other constraints that play a role in determining which agents are appropriate to punish.

satisfies the need for a method of punishment. Kantian protective deterrence replaces Kant's commitment to the traditional policy of *ius talionis* with rehabilitative means of determining the appropriate method of punishment.

Fifth and finally, a theory of punishment must offer a means of determining the appropriate **amount** of punishment.[29] This is a difficult component for any theory, as quantifying the amount of punishment poses serious challenges. Most theories of punishment aspire to articulating an amount of punishment that is equal, proportionate to, or fitting the crime, but are either uncertain or unspecific about how such proportionality is to be measured and assessed. Some theories point to the harm caused,[30] while others are more concerned with the rights that the crime violated. Despite the difficulties involved, any adequate theory must attempt to offer some fixed way to determine how much punishment is appropriate in any instance of crime. Although it is tempting to utilize rehabilitation to guide the means of specifying the appropriate amount of punishment, as I did with the method, certain asymmetries make deterrence a better standard for fixing the amount of punishment.

---

[29] The amount of punishment and the methods of punishing are very closely linked. It is impossible to specify the amount of punishment a criminal should experience if there is not some specific method by which that punishment is to be administered already in mind. Conversely, however, part of the consideration of what makes a particular method of punishing appropriate might, in some cases, be its capacity for achieving the required amount of punishment (e.g., a judge might select a fine as the appropriate method of punishing some act of crime if the details of the crime make it such that any form of imprisonment, even for a short period of time, would necessarily be too great an amount of punishment).

[30] Even determining the amount of harm caused by a given crime can be incredibly difficult, if the loss of future prospects and freedom of choice is taken into account. This matter becomes even more complex in the realm of civil law, but as punishment is not involved, we need not attempt to solve those problems here.

It is worth reiterating that although each of these elements is distinct, they are not all of equal significance for the theory. If a theory is justified by deterrent concerns, for instance, then the possible answers to questions of liability, method, and amount of punishment must all be ones that are at least consistent with the justification of the practice. There is some room for variety in the answers that can be given[31] – they need not all aim to maximize deterrence – but it cannot be the case that a deterrent justification for punishment can allow for the other elements to be ones that actively undermine or diminish the deterrent efficacy of the institution of punishment.

In exploring Kant's theory of punishment, then, this dissertation seeks to identify Kantian answers to each of these five elements. As I have described above, it is my intention to do so in a way that produces a consistent theory that respects the most foundational characteristics of Kant's practical philosophy, while also preserving as many of Kant's statements on punishment as possible.

1.4 Outline

The foregoing schematization of what a theory of punishment entails does more than just map out the conceptual space of the dissertation. In addition, this division of the elements of theories of punishment will also serve to provide the core structure of the dissertation. Aside from a few exceptions, each chapter will be devoted to exploring

[31] For instance, it is possible that a deterrent theory of punishment could specify the adoption of either retributive or rehabilitative methods of punishment if it were found that the first were more effective at deterring others from the commission of crime while the latter were better at preventing recidivism. In this way, both would be consistent with the justification for deterrence, although still governed independently – at least to some extent – by their own internal logics.

one element of Kant's theory of punishment. This will enable each of the elements to be considered separately, while still situating them within the context of a sustained examination of Kant's theory as a whole.

The next chapter of the dissertation, "A History of Violence: Punishment and the State in Early Modern Philosophy," serves to situation Kant's theory in its historical context. I explore several dominant trends in thinking about the purpose of the state and its authorization to use punitive force against its citizens. This historical analysis extends back to Hugo Grotius and includes such prominent political philosophers as John Locke, Thomas Hobbes, Samuel von Pufendorf, Jean-Jacques Burlamaqui, Cesare Beccaria, Jean-Jacques Rousseau, and Adam Smith. By examining the prevailing opinions of Kant's predecessors, I lay the foundation for the argument that Kant has foreclosed the possibility of a retributive theory by relying heavily on the traditional structure of political authority used by the natural law and social contract theorists. Although Kant strives for radical originality in his moral philosophy, his political philosophy is too heavily indebted to the liberal tradition to allow him radical free reign. Perhaps without intending to, Kant has endorsed a framework that limits the very intelligibility of a retributive theory of punishment. This traditional structure is developed in a way that almost necessitates a deterrent theory of punishment.

In the third chapter, "Defining Punishment: Coercion and Right," I articulate and defend the definition of punishment that Kant employs in his theory. Specifically, he holds punishment to be a legal institution that is impossible without determinate, publically-articulated laws and an established executive authority with the power to

enforce such laws – a 'rightful condition.' Without a rightful condition, there might be morally permissible violence that looks like punishment, but it could not be genuine punishment. In order to defend Kant's definition, I look to the division of ethical and juridical duties that he establishes in his *Metaphysics of Morals*. There is a long-standing debate between interpreters of Kant's work over the exact nature of this division; I argue that the appropriate way to understand it is by categorizing ethical duties as those that cannot be enforced, while juridical duties are those that can. This distinction helps to ground the exact nature of punishment in Kant's theory: the enforcement of juridical duties, which themselves can only exist in a state and under a rightful condition. The chapter concludes with a consideration of the role that reward could play as an incentive to perform one's juridical duties.

The fourth chapter of the dissertation, "Justifying Punishment: Deterring Threats to Freedom," takes up the question of the justification of punishment that Kant employs. In many respects, this is the core of the argument made over the course of the dissertation. Here, I examine the two arguments that Kant gives in favor of a retributive justification for punishment. I reject the first as untenable and at odds with the very definition that Kant has established. The second argument works, but functions only as a negative constraint rather than as a positive reason for adopting retributivism. If another kind of theory could satisfy this constraint, then Kant could give no reason for thinking that this alternative justification is impermissible. I attempt to show how a deterrent theory could satisfy this constraint in the next chapter. This chapter also explores and rejects the dominant version of Kantian deterrence – as represented by Byrd and

Ripstein – for being too state-focused. Instead, I propose Kantian protective deterrence as a more intuitive, individual-focused deterrent justification for punishment.

The second half of chapter four provides the positive argument for Kantian protective deterrence. The goal is to show how Kant's own fundamental practical principles support a deterrent approach to justifying the state's use of coercive force against its own citizens. By looking to the 'universal principle of right,' we can develop an account of how Kantian deterrence might work. This account relies heavily on the 'hindering of a hindrance to freedom' argument and on the purpose of the state.

In the fifth chapter, "The Liability of Punishment: All and Only Those Who Have Done Wrong," I turn to the question of how Kant's theory specifies the appropriate targets or recipients of punishment. In particular, I take the challenge for a Kantian deterrence theorist to be explaining how a deterrence theory can still explain why we ought to punish all and only those who have committed crime. I offer two arguments for this limitation, one practical and one moral. I also demonstrate how many of Kant's most retributive-sounding passages can be accommodated within a deterrence theory by structuring the liability in a broadly retributive manner.

The dissertation's sixth chapter, "The Amount and Method of Punishment: *Ius Talionis* and the Formula of Humanity," focuses on the appropriate amount and methods of punishment that Kant's theory endorses. These two elements of the theory are addressed together for several important reasons. First, it is difficult to conceptualize either the amount of method of punishment without making reference to the other. Second, Kant offers the same basis for selecting the appropriate amount and methods of

punishment: *ius talionis*, the law of retribution. Kant holds that such determination must be made in accordance with as literal an equivalence as possible, writing,

> But what kind and what amount of punishment is it that public justice makes its principle and measure? None other than the principle of equality…. Accordingly, whatever undeserved evil you inflict upon another within the people, that you inflict upon yourself. (*MoM* 6:332)

This principle is expressed familiarly as punishing 'an eye for an eye.' In the chapter, I will argue that Kant's adoption of *ius talionis* is deeply flawed, both in ways that he recognizes and in ways that he does not. For instance, there are a great many forms of crime for which there can be no equivalent in form or amount, for both moral and practical reasons. In all of these cases, *ius talionis* fails us.

Instead of *ius talionis*, I contend, Kant should look to his own fundamental moral principles for guidance on the appropriate amount and methods of punishment to employ in his theory. Specifically, the second formulation of the categorical imperative – the Formula of Humanity as an End, or FHE – is a better guide for Kant than *ius talionis*. This substitution would help Kant to capture many of the concerns that are underlying his embrace of retributivism – for instance, his fear that other forms of punishment use persons as a means to some end, like deterrence – but do so in a way that does not create just as many problems as it solves.

In the seventh chapter of the dissertation, "*Quis Custodiet Ipsos Custodes*?: Revolution and Punishing Rulers," I consider some of the implications of Kant's theory of punishment. In particular, I focus on questions of the permissibility of civil disobedience, revolution, punishing former leaders of a state, and international criminal courts. In the case of civil disobedience and revolution, Kant strongly rejects the legality

or morality of ever engaging in any such behaviors. The way in which he defines the sovereignty of the legislative branch and the authority to punish of the executive branch frames the issue in such a way that claiming a right for citizens to resist the power to the state is potentially incoherent. I argue, however, that while this may be the case for a legal right to revolution, Kant oversteps the limits of position when he claims that resistance is always immoral. Indeed, Kant's own moral philosophy should incline him toward thinking that there is a special class of cases in which resistance is not only morally permissible, but indeed required of the citizens. With respect to the punishment of former leaders and the viability of international criminal courts, Kant's position is slightly more complex. I analyze these issues and consider what conclusion we ought to draw, basing my analysis on his work *Toward Perpetual Peace*.

As becomes clear from the above outline, this project approaches Kant's views on punishment from a number of different angles and vantage points. By utilizing this broad analytical lens, I hope to offer the most comprehensive account of Kant's position on punishment to date. This exhaustive scope is not only useful for ensuring that no important details are overlooked, but also for guaranteeing that Kantian protective deterrence is consistent with Kant's broader practical philosophy.

2 A History of Violence: Punishment and the State in Early Modern Europe

In his practical philosophy, just as in his more speculative work, Kant strove to articulate a substantively innovative solution to the philosophical questions he addressed. His efforts to achieve originality of perspective, however, were not founded on an ignorance of or disregard for the work of previous philosophers; if anything, some of his characteristic innovations were envisioned as a synthesis of established, prevailing views. Any attempt to understand Kant in isolation, then, is able to capture only a part of the full picture. To truly understand his work, we need an understanding of his influences and the major contributors to the on-going discussion of his day and age.

This is especially true of Kant's political philosophy. While his work on moral philosophy is framed largely by his opposition to previous ways of thinking, his positions in political philosophy draw heavily from the work of major figures in the history of western political thought. Although he introduces his own distinctive elements, there is no denying that his writings on civil society, the state, sovereignty, and law are inspired by the natural law, social contract, and British moral sense traditions that preceded them.

In this chapter, I will be exploring the work of a number of political philosophers from the 17th and 18th centuries. Not only are these all figures with whom Kant is known to have been familiar, but they also all exerted great influence on the general conversation about political authority and punishment and therefore deserve

consideration independently of Kant's acquaintance with their work. My primary aim is to determine each philosopher's position on two matters: 1) the relationship between the state and punishment, and 2) the justification that each thinker offers for the state's use of coercive force against its own citizens. Where possible, I will also pursue a secondary goal of outlining what, if any, specifics are provided or constraints are imposed to specify the appropriate recipients, amount, and method of punishment. Clearly, this will require careful consideration of the passages in which each thinker describes the sovereign's executive powers and the state's use of punishment. In addition, however, we must look at the views that the philosophers hold regarding the purpose and appropriate role of the state itself, as this often serves as an indication of what limits or constraints they place on punishment. If the state exists for a given reason, it is often the case the punishment will be justified in light of the same reason.

The conversation about the definition of punishment – that is to say, what criteria an act of violence must meet in order to be punishment – resolves into two broad categories. First, there are those that define punishment as a kind of natural right. According to philosophers who hold this view, punishment – and a right to punish – exists prior to or independently of states and human laws. Furthermore, any acts of punishment that do occur within civil society are connected to and based upon the right to punish that exists in the state of nature. Put another way, this category of thinkers holds that humans are capable of punishing one another for reasons that do not originate in the civil authority of the state. Given this claim, all members of this group justify punishment without reference to the role it plays in civil society or lawful states.

Although no one would deny that punishment can and does serve useful functions, they think that these functions are only possible given that punishment has already been found permissible. Additionally, they justify the state's authorization to punish – that is, to utilize force against its citizens in the execution of its laws – by grounding such an authorization on these independent, pre-state considerations. Kant does not endorse the unifying claim defended by this school of thinkers, but it is worth exploring their line of thought in order to get a sense of the major disputes of the era.

Rather, Kant follows in the tradition of the second main category: those who hold the state's right to punish to be different in kind from – and unconnected to – any act of violence that can take place in the state of nature. Some, like Kant, hold punishment[1] to exist only in states with established law and a recognized executive that has authority to enforce the laws. Others might think that some form of punishment exists in the state of nature, but that this form and the one perpetuated by states are fundamentally different; the state's authorization does not depend on the pre-civil right to punish. Within this broad category, I will be distinguishing between two important sub-groups. The first, to whom I will refer as the 'strict contractarians,' is committed to the idea that the state's purpose is derived not from any prior normative facts about human kind, its natural rights or duties, or its ends, but rather from the act of contracting the state itself. The

[1] Please note that when I refer to punishment, I mean *human* punishment. There are those who, despite viewing human punishment as a construction of states, nevertheless think that God might be able to punish humans in the state of nature. Arguably, this does not change their overall stance, as God punishes with the authority of a sovereign and on the basis of laws, simply not laws of a state constructed by humankind. I will not be considering this view in any great depth, as our focus here is on punishment by human beings and governments composed of human beings.

strict contractarians take it to be the case that people enter civil society for reasons of rational self-interest, and this same rational self-interest can explain what kind of state the contractors would agree to enter. The second sub-group, to whom I will refer as the 'normativists,' holds that the existence of the state is not merely a neutral occurrence that arises from rational self-interest. Rather, the state is seen in a normative way; either citizens have some moral duty to enter the state, or else it is necessary for helping them achieve some end that carries normative force. As such, we do not derive the purpose of the state from the rational self-interest of the contractors, but rather from the normative reasons for the state's existence. It is to this last sub-group that Kant's political works belong.[2]

Throughout this chapter, I will repeatedly make the case that either explicitly or implicitly, almost all political philosophers during the early modern period – regardless of their association with either the natural right tradition, the strict contractarian tradition, or the normativist tradition – held theories of punishment that, while mixed to various degrees, nevertheless utilize an ultimately deterrent justification for punishment. A truly, deeply retributive theory of punishment of the sort that has historically been ascribed to Kant would have been a dramatic departure from the broad consensus of the age. Furthermore, in all cases the adoption of a deterrent view was not a matter of mere coincidence; rather, it was necessitated by the details of the position's

[2] Note that these categories and sub-groups do not represent any kind of official affiliations or regimented schools of thought that existed during the 17th and 18th centuries. While the thinkers in question make clear their support for the views that I utilize as unifying principles for the various groups, these distinctions are solely meant to draw out important differences in the ways in which early modern political philosophers approached punishment and its connection to the state's purpose.

answer to questions regarding the origin of punishment and the purpose of the state. Put simply, the common views of the ends of civil society in 17th and 18th century Europe required widespread consensus on the deterrent grounds for punishment.

2.1 Punishment as a Natural Right

Throughout the early modern period, a number of philosophers and political theorists endorsed the idea that punishment was not merely an invention of states, but rather a natural right belonging to human beings independently of their association with any state or civil society. In one way or another, each of these thinkers based this claim on the idea of natural law or laws of nature; such laws exist independently of states, their existence enables punishing occur outside of state institutions, and their violation—or the prevention thereof—serves as the justification for punishment. By looking at three of these philosophers individually, one can get a sense for what elements were shared by members of this tradition. Specifically, the works of Grotius, Locke, and Burlamaqui, although different in many respects, grant to humans a natural right to punish on the basis of transgressions of the law of nature, legislated by God and knowable through the exercise of human reason. These laws, in turn, were authored with the specific goal of improving the lives of human beings, securing our well-being, or leading to our highest individual happiness, and as such our interest in following and promoting the observance of these laws is both a duty to God and a matter of self-interested prudence.

Grotius

Hugo Grotius is frequently taken as the starting point of works on natural law theory in the modern period. There is a rationale behind this approach: his work takes steps away from a strict reliance on revelation with an eye toward establishing a secular basis for natural law. Indeed, he is cited by later natural law theorists such as Pufendorf and Barbeyrac as the father of modern natural law theory.[3] These steps aside, however, Grotius still shares much with the medieval way of approaching philosophy. A quick perusal of his masterpiece *The Rights of War and Peace* shows his continued reliance on the scholastic habit of appealing to authority. There has been a long-standing debate regarding the degree to which his work is secular,[4] in light of the view that his traditional, conservative form of writing merely allows Grotius to progress toward his original, secular goal with less opposition from the religious and intellectual establishment of the 17th century. Whatever the case may be, his efforts to incorporate the received wisdom of centuries of jurisprudence lead to a certain amount of tension in his work. Despite the pains he takes to weave together biblical citations with passages from Greek philosophers, Roman jurists, and medieval scholastics, unsurprisingly he cannot fully avoid some conflicts. Among the other areas of tension, Grotius's position on punishment tries to unite several different ways of thinking without a clear description of how these disparate views are to be reconciled. A charitable reading of the

---

[3] Irwin, Terence. *The Development of Ethics: Volume II: From Suarez to Rousseau*. Oxford: Oxford University Press, 2008. p. 322.

[4] Haskell, John D. "Hugo Grotius in the Contemporary Memory of International Law: Secularism, Liberalism, and the Politics of Restatement and Denial." *Emory International Law Review*, Vol. 25, No. 1 (2011).

end result reveals a kind of mixed theory, albeit one that does not fully explain the way in which the mixed elements interact with one another. What is clear, however, is his commitment to the view that punishment exists in the state of nature as a natural right.

Before addressing the issues that he faces with justification, let us turn to Grotius's discussion of the definition of punishment. His first, most general statement on the subject of punishment is too broad to indicate much about his theory. He writes, "Punishment taken in its most general meaning signifies the pain of suffering, which is inflicted for evil actions."[5] There are certain retributive elements hinted at by the final clause, and the description of punishment as "pain of suffering" might tell us something about the method of punishment, but we are nevertheless not much closer to understanding what kinds of acts of violence are counted as punishment and why such acts are justified.

We can get a better sense of what Grotius takes to be the nature of punishment by examining what he says with respect to who has the right to punish wrongdoers. Grotius is clear that punishment can only be inflicted by one who has a right;[6] violence by one who lacks a right to punish, even if it is directed at one who has done wrong on the basis of her or his wrongdoing, cannot be rightful punishment. As for who he takes to have such a right to punish, the matter is more complicated. He holds that the question of who has the right to punish offenders is not one determined by a fixed law of nature (although the necessity of punishment is itself determined by the law of

[5] Grotius, Hugo. *The Rights of War and Peace*. Washington: M. Walter Dunne, 1901. p. 221.
[6] p. 222

nature).[7] He lays out all that we can gain from a well-reasoned consideration of the issue:

> It is deemed most suitable for a superior only to be invested with the power of inflicting punishment. Yet this demonstration does not amount to an absolute necessity, unless the word superior be taken in a sense implying, that the commission of a crime makes the offender inferior to every one of his own species, by his having degraded himself from the rank of men to that of the brutes, which are in subjection to man.[8]

In other words, the violation of laws of nature renders one deserving of punishment – or, at least, demonstrates that one so deserves. Reason, our source of knowledge about the laws of nature, cannot tell us what specific person or persons are authorized to punish such violators. Grotius takes it as a matter of self-evident rationality, though, that the punisher ought to be a superior. The only way to make sense of this in the state of nature, he argues, is to take what will later become the standard natural law position: by violating the law of nature, the violator proves to be of an inferior sort, like an animal or a 'brute.' A human might not have the right to engage in violence against her or his equal, but there is no question in Grotius's mind that she or he does possess such a right against animals. If wrongdoers are like animals in their inability to act in accordance with law, then they are also vulnerable to abuse at the hands of other humans. This means that all law-abiding persons, insofar as they are human, are now naturally superiors and can therefore punish the wrongdoer.

[7] p. 223
[8] Ibid.

This account of punishment in the state of nature is well supported by what he has to say on the subject of war between states.[9] As Grotius's focus was on establishing rightful conditions of war that would, in turn, allow for peace to exist between nations, much of his book is given over to considerations of the various kinds of justifiable conflict. One form of international violence that he takes to be potentially justifiable is war motivated by an interest in punishing another nation for inappropriate actions. The relationship between nations is parallel to the situation of persons in the state of nature, meaning that if punishment is possible in the former case, it would also be possible in the latter. Presumably, the rationale is also similar; if a nation engages in behavior that gives others a claim of superiority, then it is punishable by these other states.[10]

If punishment is warranted in the state of nature due to the inferiority of one who would violate a natural law, what kind of basis can there be for punishment in civil society? After all, Grotius does not think that just anyone in a state has the right to punish violators of the state's laws. This power belongs exclusively to the sovereign.[11] The reason given for the sovereign's exclusive claim to the use of force in executing the laws and punishing violators also has to do with superiority. In this case, the sovereign

---

[9] Ibid., p. 75

[10] Although Grotius does not discuss limitations on punishments that can be inflicted against individuals who violate the laws of nature, we might find evidence that he envisioned such limitations by considering what he says on the subject of war. Given the parallel between states and individuals in the state of nature, the fact the he argues in favor of limiting what acts a state can engage in during warfare (for instance, he prohibits), would suggest that he favors similar kinds of limitations on individual punishment in the state of nature.

[11] Grotius allows for the possibility of a division of powers, in which case there might be a designated executive authority who is distinct from the person or body that holds supreme sovereign authority. In such a case, however, the executive is still merely an agent of the sovereign; any power that he or she wields is itself legitimate only because it is a function of the sovereign's power.

is superior by virtue of her or his position in two senses. On the one hand, the members of the state have agreed to give such power to the sovereign. Grotius believes the governed give their consent, thus imbuing the sovereign with all the authority that he or she needs. On the other hand, the sovereign, as the one who creates laws, also has the power to enforce the laws—the same way that God is the only one who automatically enjoys the kind of superiority necessary for enforcing the laws of nature.

Grotius explicitly connects the right to punish that all persons outside of the state enjoy with respect to violators of the law of nature with the right to punish that the sovereign has with respect to the citizens of her or his state. The latter power is merely the extension of the former into a new situation. There are not two kinds of punishment, but rather a single kind that can occur in rather starkly different situations. It is reasonable, then, to expect that all punishment – whether in the state of nature or civil society – would be justified in the same way.

What kind of justification is this, then? Unfortunately, this is where *The Rights of War and Peace* is weighed down by the amount of extant "authority" with which it must harmonize. Indeed, there are elements of all three of the traditional triumvirate of possible grounds for punishment: retribution, deterrence, and rehabilitation. While there is nothing inherently contradictory about trying to incorporate elements of these three different views of punishment into a single position, Grotius does not give us a perfectly picture of how they are meant to fit together.

He indicates an affinity for retributive theories when he claims that "it is right for everyone to suffer evil proportioned to that which he has done,"[12] or "when we say that punishment is due to anyone, we mean nothing more than it is right he should be punished."[13] In these and related instances, he defends the kind of analyticity of the relationship between wrongdoing and punishment that is characteristic of retributive theorists. Whether these retributive concerns are sufficient to justify punishment on their own, however, remains unclear.

There is good reason, I believe, to think that the view he is proposing is not merely retributive. To start with, he writes that the mentality that "the pain of an enemy, considered solely as such, is no benefit to us, but a false and imaginary one."[14] This seems to undermine his former retributive inclinations, or at least cast them in a different light. While moral desert is essential for justified punishing, it is not alone sufficient to justify such action. To be motivated solely by an interest in causing suffering, however deserved, is to pursue a "false and imaginary" benefit.

Grotius also argues that in some cases, punishment can be mitigated for reasons of mercy.[15] The possibility of clemency does not directly contradict the use of a retributive justification for punishment, but it does complicate the picture. It is not clear if mercy truly is the motivation for clemency, though, given that he then states that this mercy is often the most effective means of causing a criminal to reevaluate her or his

[12] Ibid., p. 221
[13] Ibid., p. 222
[14] Ibid., p. 225
[15] Ibid., p. 224

actions and refrain from future crime. While I think that it is possible to understand this line of thinking as ultimately aimed at deterrent goals – and will give an argument to that effect below – there is also a decidedly rehabilitative sound to his claims.

These are not the only rehabilitative elements in Grotius's theory, though; he goes so far as to describe rehabilitation of the criminal as a justifiable end of punishment. "The power of inflicting the punishment, subservient to this end [rehabilitation], is allowed by the law of nature to any one of competent judgment, and not implicated in similar or equal offences."[16] He goes on to add that in civil society, we must establish particular individuals as competent judges, but it is worth noting that in this quotation we see an additional instance of Grotius stating that punishment is possible in the state of nature, carried out by one who has not committed a similar offense and is thus a superior, independent of laws or an established sovereign with executive authority.

Alongside these rehabilitative (and retributive) elements, Grotius also makes claims that could easily be viewed as grounding punishment on deterrence. He writes, "In punishments, we must either have the good of the criminal in view, or that advantage of him whose interest it was that the crime should not have been committed, or the good of all indifferently."

With enough effort, we might find a way to render these statements all consistent with one another, most likely by subsuming some of his retributive and rehabilitative interests under a general drive for deterrence. Likewise, all of the other elements of

[16] Ibid., p. 226

Grotius's theory of punishment could conceivably be incorporated into a coherent whole. Like every other philosopher described in this chapter, Grotius is concerned with establishing a certain kind of equivalence between punishments and the crimes to which they respond. The exact nature of this correspondence is left ambiguous, however, and I will not take the time here to try to sort out what kind of equivalence he might have in mind.

Instead, I will conclude this section by focusing on the lasting impact that Grotius's way of approaching the issue of punishment had on early modern political philosophy. By connecting the state's authority – and its right to punish – with an individual, natural right to enforce the normatively significant laws of nature, Grotius was able to provide an explanation for the state's legitimate authority that did not rely on divine selection, shaky metaphors to paternal power, or simple might. Although some later thinkers would rely solely on the social contract and the consent (actual or hypothetical) of the governed, Grotius's introduction of rights, knowable through reason and normatively powerful, served as a basis for at least some later political philosophers. As well will see, Locke and Burlamaqui both relied heavily on the kind of secular option that Grotius first posited.

Locke

Although Grotius is the first philosopher of the early modern period to make the case for punishment as a natural right, this position is perhaps most commonly associated with the political philosophy of John Locke, as laid out in his *Two Treatises of*

Government. Indeed, the possibility – and even requirement – of extra-state punishment is one of the most distinctive features of Locke's conception of the state of nature. In order to assess the other features of Locke's theory, we must know why he holds punishment to be possible in the state of nature and why he argues in favor of viewing it as a universal right.

Let us turn to what Locke has to say on the subject of punishing. For Locke, the subject comes up much earlier than it does for most other social contract theorist or natural law thinkers. In describing humanity's situation in the state of nature and what rights individuals hold in such a condition, he writes:

> But though this be a *state of liberty*, yet *it is not a state of license*...The state of nature has a law to govern it, which obliges every one: and reason, which is that law, teaches all mankind, who will but consult it, that being all *equal and independent*, no one ought to harm another in his life, liberty, or possessions.
>
> And that all men may be restrained from invading others' rights, and from doing hurt to one another, and the law of nature be observed, which willeth the peace and *preservation of all mankind*, the execution of the law of nature is, in that state, put into every man's hand, whereby every one has a right to punish the transgressors of that law.[17]

Unlike his predecessor Hobbes (who I discuss below), Locke does not take the state of nature to be morally neutral. Instead, it is given a normative character by the existence of the laws of nature, which are discernible through the use of reason. There is some ambiguity in the passage above about the origins of these laws; given his language, we might be tempted to conclude that the laws are merely a product of

[17] Locke, John. *Second Treatise of Government*. 1690. Ed. C.B. Macpherson. Indianapolis: Hackett Publishing Company, Inc., 1980. Pp. 9-10.

human reason. To conclude this, however, would be to ignore other important quotations that offer support for a different view. I will return to this point shortly.

One other important point in the passage quoted above that warrants consideration is his claim that all people hold this executive right to punish. This leads to one of Locke's more unique arguments for leaving the state of nature: the practical necessity of consolidating this executive right of punishment.[18] Instead of everyone attempting to punish violators of the laws of nature individually (and potentially allowing personal feelings to illegitimately influence the amount of punishment applied), civil society allows for the designation of a single individual or group as responsible for any and all acts of punishment that need to occur. While this is not the only reason Locke gives for individuals' move from the state of nature to an established and lawful state, the fact that he includes it only reinforces the centrality of punishment in his political philosophy.

Like Grotius, Locke thinks that the sovereign's ability to punish in a state is an extension of this executive right in the state of nature. He conceives of the state as limited in its possible powers to those that the individuals who comprise the state had themselves prior to joining the state.[19] There is no other source of state power or authority, and so any ability to punish must be based on the individual right that exists in the state of nature. Rather than concluding that the omnilateral, reciprocal agreement of all members is what gives rise to the that state's authorization to punish, as some later

[18] Ibid., p. 12
[19] Ibid., p. 70

social contract theorists posit, Locke holds that this agreement is merely a consensus to allow a specific individual or body to be the sole party to act on the universal executive right that all possess.

Moving on from the definition of punishment that Locke offers, let us consider the justification he offers for this universal executive right to punish enjoyed by all people in the state of nature. On the surface, it is clearly a direct response to an act that violates the laws of nature. Violence against one who has not violated the laws of nature cannot be punishment, and the mere commission of an act that contradicts the law of nature serves as appropriate grounds for punishment. After a quick reading, then, this seems to be an open and shut case of retributive thinking.

Before we conclude this, though, we should return to the question I set aside earlier: namely, from where do the laws of nature come? Locke is clear that these laws are the product of God's will.[20] In order to avoid a contradiction, then, we should not take his claim that reason is the law of nature to mean that the prescriptions contained in the law are merely the products of reason. Rather, reason is a necessary and sufficient tool for understanding the contents of the laws of nature, which are given by God. He writes, "In transgressing the law of nature, the offender declares himself to live by another rule than that of reason and common equity, which is that measure God has set to the actions of men, for the mutual security."[21]

[20] Ibid., pp. 10-13
[21] Ibid., p. 10

This last quotation is important. The laws of nature, given by God and intelligible through reason, are not arbitrary. Rather, they have a specific set of tangible goals that they are meant to produce: human security. When describing the motivation for punishing those who have violated the law of nature, Locke's focus on the purpose of the laws of nature comes out very clearly:

> Every man, in the state of nature, has a power to kill a murderer, both *to deter* others from doing the like injury…by the example of the punishment that attends it from every body, and also to secure men from the attempts of a criminal, who having renounced reason, the common rule and measure God hath given to mankind, hath, by the unjust violence and slaughter he hath committed upon one, declared war against all mankind, and therefore may be destroyed as a *lion* or a *tyger*, one of those wild savage beasts, with whom men can have no society nor security.[22]

Despite any initial temptations to view Locke's natural, universal right of punishment as grounded in a retributive manner, it is clear from what he says above that he views punishment as justified for deterrent reasons. No mention is made of desert, but rather only of security and making an example so that others might learn vicariously. Even the motivation he describes is not one that sounds classically retributive; we punish criminals as we would wild animals, not because they deserve it, but because it is the only way to prevent them from future acts of wrongdoing.

This also gives us a clear answer on the question of what Locke takes to be the proper liability of punishment. We should punish those who have done wrong, as they are the ones who have demonstrated that they cannot be trusted in the future. As a secondary benefit to the general deterrence of crime, others who have yet to prove

[22] Ibid., p. 11

themselves untrustworthy may still be convinced to follow the law by witnessing what happens to those who do not. In this way, Locke fixes the recipients of punishment retributively, as we have seen is common even among those with a deterrent justification for the government's use of force against its citizens. This kind of mixing of the elements was prevalent among early modern jurists precisely because it enabled their deterrent theories to still capture some of the strengths of retributive and rehabilitative theories. As we will see, this same kind of mixing represents the best way for Kant to solve some of his difficulties with punishment.

As for other aspects of Locke's theory of punishment, he lays out his position on the appropriate amount in the following way: "I answer, each transgression may be *punished* to that *degree*, and with so much *severity*, as will suffice to make it an ill bargain to the offender, give him cause to repent, and terrify others from doing the like."[23] Unsurprisingly, given his commitment to a deterrent justification for his theory of punishment, Locke suggests a way of fixing the amount of punishment that relies entirely on deterrent reasoning. Again, this understanding of the appropriate amount of punishment was cited more frequently than any other during the early modern period, and we have already seen other thinkers who hold something quite similar.

The last important consideration in the amount of punishment that Locke holds to be justified also provides some insight into his limited thoughts on the appropriate method of punishment. Although he does not have much in the way of a positive view

[23] Ibid., p. 12

of what particular forms of punishment are best or appropriate, he does consider objections to one special kind of punishment: death. Given that the appropriate amount of punishment is fixed in accordance with what is necessary to have deterrent effect, it is likely that lesser punishments can often satisfactorily deter crime, meaning the death is often unnecessary.[24] Nevertheless, should it be the case that death is what is needed, Locke has no objections to its use as a form of punishment.

In many ways, Locke best represents the great influence and deep roots of the natural law position on punishment. Despite his affiliation with the social contract tradition, his commitment to the normative laws of nature ultimately leaves him with a view that is remarkably similar to Grotius's. By accepting the basic framework for justifying punishment provided by Grotius, Locke ultimately functions as one of the clearest examples of a kind of deterrent justification for punishing criminals, in light of their inability to live in accordance with the rationally intelligible laws of nature.

Burlamaqui

Burlamaqui does not have much of a reputation among historians of philosophy as an innovative philosopher, jurist, or natural law thinker. His work is often cited as having had a substantial readership among the framers of the Constitution of the United States, but was nevertheless not known for originality. Instead, he is frequently seen as borrowing heavily from Barbeyrac, the noted commenter on and popularizer of

[24] Ibid.

Pufendorf's work. It is clear from his work *The Principles of Natural and Politic Law* that he holds Barbeyrac in the highest regard. In at least one respect, however, Burlamaqui departs from the Pufendorf-as-interpreted-by-Barbeyrac orthodoxy: namely, he introduces a strong positive and teleological component to natural law. Rather than merely preventing humans from harming one another as his predecessors in the natural law tradition had held and as Locke had argued, natural law is, according to Burlamaqui, imposed by God to help human beings achieve their natural end of happiness. It justifies punishment in the state of nature, and it leads the state to adopt practices and institutions of punishment that are structured and grounded in a deterrent manner.

Burlamaqui's stance on the definition of punishment is not particularly complicated, but it is worth working through how he sets it up. He offers a clear enumeration of what he takes to be punishment's necessary features, but he notes that his definition only applies to punishment in civil society. How he relates this technically precise definition to the possibility of punishment outside of civil society lays the foundation for his justification as well. I will first lay out the definition he gives, followed by sketching out how this understanding of state punishment is grounded on a right to punish in the state of nature.

Punishment, according to Burlamaqui, is "an evil, with which the prince threatens those who are disposed to violate his laws, and which he really inflicts, in a just proportion, whenever they violated them, independently of the reparation of the damage, with a view to some future good, and finally for the safety and peace of

society."[25] Burlamaqui covers quite a bit in this admirably clear sentence, touching on everything from the definition to the appropriate liability and amount of punishment. Given the space constraints of this chapter, I will be focusing exclusively on the issues raised in the second and third clause: namely, that the prince is the one responsible for threatening and inflicting punishments on those who violate the laws.

This focus on the sovereign as the only one who is capable of punishing, as well as the description of those who are punished as those "who are disposed to violate laws," might give the impression that Burlamaqui should properly be understood as holding punishment to be a construction of states, impossible without a sovereign and laws. He makes it clear, however, that this is not the case. "Not that every punishment in general supposes sovereignty, but because we are here speaking of the right of punishing in society, and as a branch of supreme power."[26] This statement, then, indicates that there is some form of punishment which can occur without sovereignty and outside of society.

Just what would this extra-social form of punishment look like? It is essentially the same as Grotius's and Locke's accounts:

> Whoever violates the laws of nature, testifies thereby, that he tramples on the maxims of reason and equity, which God has prescribed for the common safety; and thus he becomes an enemy of mankind. Since therefore every man has an incontestable right to take care of his own preservation and that of society, he may, without doubt, inflict on such a person punishments capable of producing repentance in him, of hindering him from committing the like crimes for the future, and even of deterring others by his example.[27]

---

[25] Burlamaqui, Jean-Jacques. *The Principles of Natural and Political Law*. Indianapolis: Liberty Fund, 2006. p. 418
[26] Ibid.
[27] Ibid., p. 417

Punishment in the state of nature can be inflicted by anyone. While I will consider the specifics of why such a universal right to punish is justified shortly, for now it suffices to say that it is the violation of natural law that opens one up to punishment from her or his fellow humans. Another point worth mentioning is that Burlamaqui takes this right to inflict coercive force against the perpetrators of crimes against the laws of nature to be the same as the right of war.[28]

So, we have seen then that Burlamaqui holds punishment to be possible in the state of nature, as a right held by all against those who violate the laws of nature. He also describes it as occurring in states, under much more limited and constrained circumstances. What is the relationship between these two kinds of punishment? He answers this question clearly:

> The right of executing the laws of nature, and of punishing those who violate them, belongs originally to society in general, and to each individual in particular; otherwise the laws which nature and reason impose on man, would be entirely useless in a state of nature, if no body [sic] had the power of putting them in execution, or of punishing the violation of them….By following these principles, it is easy to comprehend that the right of a sovereign, to punish crimes, is no other than that natural right which human society and every individual had originally to execute the law of nature, and to take care of their own safety; this natural right has been yielded and transferred to the sovereign, who, by means of the authority with which he is invested, exercises it in such a manner, as it is difficult for wicked men to evade it.[29]

As with Grotius and Locke, the right of the state to implement a system of punishment and carry out specific coercive acts against its citizens is based on natural

[28] Ibid., p. 452
[29] Ibid., p. 417

rights held by individuals in the state of nature. The sovereign may use force for the same reasons that individuals can outside of civil society. The only question that remains, then, is whether this universal right to punish in the state of nature, and the executive right it gives rise to in civil society, is justified in a retributive or deterrent manner.

Like the other natural rights theorists, it is perfectly plausible that Burlamaqui's view – which holds punishment to be warranted a response to the violation of natural law – could be retributive; the proponents of this theory would need merely to argue that the natural law, as God's will, is of such a nature that its violation makes one deserve to suffer; the state, in turn, has a moral imperative to bring about this end. Like Grotius and Locke, however, Burlamaqui does not go down this road, choosing instead to depict punishment – both in the state of nature and in civil society – as justified for deterrent reasons.

The introduction of happiness as the end of human life has some major consequences for the ways in which Burlamaqui approaches the social contract and the state – and therefore for the way in which he justifies the state's right to punish. As we will see, the reason given by the strict contractarians for the contractors willingness to give up some liberty and join a state was essentially security. For Burlamaqui, though, the contractors are not interested solely in the freedom from violence and theft that civil society affords, but rather they see a state as a necessary tool for maximizing their own,

personal happiness.[30] He sums this up in the following way: "Civil society is nothing more than the union of a multitude of people, who agree to live in subjection to a sovereign, in order to find, through his protection and care, the happiness to which they naturally aspire."[31]

If the citizens of a state have entered it with a positive goal in addition to protection from harm, then the state must be viewed as having the purpose of bringing about this goal. Whatever the state does, then, should be aimed at increasing the individual happiness of its citizens. By including his egoistic view of human nature and motivation, Burlamaqui precludes the possibility of the state acting to promote general happiness in a highly unequal way; given that they are motivated to maximize only their own happiness, they would not consent to policies that regularly sacrifice individual happiness for the good of the general public. Some inequality would no doubt be compatible with egoistic contractors, but Burlamaqui gives us no indication that he has an opinion on what quantity of inequality would be the maximal acceptable amount.

Since the positive end of promoting happiness motivates the state's policies, it is not surprising that Burlamaqui justifies the state's authorization to use coercive force against its citizens in generally deterrent language. After all, the prevention of crime

[30] We should note that Burlamaqui holds human motivation to be strictly egoistic. There is no interest in the good of others or in happiness in general. This allies him with Hobbes, but puts him at odds with Pufendorf and Beccaria, whose consequentialism takes more of a utilitarian character.
[31] Ibid., p. 276

before it happens does more to ensure and guarantee happiness than does the punishment of criminals after the crime has occurred. He writes,

> The principle end of punishment is therefore the welfare of society; but as there may be different means of arriving at this end, according to different circumstances, the sovereign also, in inflicting punishments, proposes different and particular views, ever subordinate, and all finally reducible to the principal end above-mentioned.[32]

There is little question, then, that Burlamaqui follows the general trend of endorsing a deterrent view of punishment's purpose and justification. Despite the differences between his work and that of Grotius and Locke, he nevertheless proposes a very similar picture of punishment: a natural right of all persons in the state of nature, possible in civil society only in light of this natural right, and justified in a broadly deterrent manner.

2.2 Punishment as a State Construction: The Strict Contractarians

As we have seen, early modern political philosophy saw a strong tradition of natural law theorists asserting the existence and propriety of punishment outside of, or at least independently of, the state and its laws. In addition, the thinkers of this tradition were prone to positing specific, normative purposes of these laws, originating in the author of these laws—God himself. At the same time, however, other philosophers proposed a different understanding of the origins of punishment. In particular, they saw punishment as a construction of states; as a result, punishment by definition became

[32] Ibid., p. 421

impossible without codified laws and an established authority whose job it was to enforce these laws.

Within this broad group of philosophers who saw punishment as a state construction, I will be drawing a significant division. On the one hand are those who view all of the purposes, ends, and features of states – including the institution of punishment – to be derived entirely from the social contract and the rational self-interest of the contractors themselves. On the other hand are those who hold the state to have some normative role, and, as such, it has specific ends that any ideal state must meet. In the following section, I will be focusing on the first group: the strict contractarians.

Pufendorf

The first example of the strict contractarian view that I will be discussing is Samuel von Pufendorf. Pufendorf was an extremely prominent and influential jurist and political philosopher whose work in the natural law tradition directly inspired much of the work of later natural law and social contract theorists. Although his work *On the Duty of Man and Citizen* slightly postdates Hobbes's masterpiece *Leviathan* –the next text I will be discussing – I think there is good reason to consider Pufendorf first: namely, his views are more clearly in line with traditional ways of thinking about natural law and the state. While Hobbes's views are undoubtedly revolutionary in a number of ways, Pufendorf's more closely resemble those I analyzed above. His break from the view that punishment is a natural right, therefore, is decidedly more nuanced and can be seen as symbolically linking the positions of Grotius and Hobbes.

Pufendorf is clear that the state of nature is not a morally neutral world. Rather, it is one in which humans are still bound by a great number of obligations. These obligations can be directed at oneself, one's fellow humans, or God, but all obligations in the state of nature, regardless of to whom they are directed, originate from God. This is owing to Pufendorf's claim that obligation requires a superior who has both just authority over us and the power to punish when that authority is ignored.[33] In the state of nature, no one has either of these necessary powers over us other than God. As such, punishment of human beings by human beings is impossible. Although it is likely that individuals would still engage in coercive or violent actions with the goal of 'punishing' the target of their actions, these behaviors would not constitute proper, justified punishment.

It is important to note, however, that among the obligations humans have in the state of nature, there is no duty to enter into civil society or to found a state. True, Pufendorf thinks that there are very good reasons for us to enter a state; without it, we are helpless and powerless to prevent others from violating their oaths and betraying us.[34] These prudential reasons, however, are the only ones he offers for the reason why persons or communities give up their natural liberty and join into lawful states. Remaining in the state of nature, while foolish, would not be in violation of a moral duty. He writes,

> It is not enough to say here that man is drawn to civil society [*societas civilus*] by nature herself, so that he cannot and will not live without it. For man is

[33] Pufendorf, Samuel von. *On the Duty of Man and Citizen*. Ed. James Tully. Cambridge: Cambridge University Press, 1991. Pp. 27-28.
[34] Ibid., p. 130

> obviously an animal that loves himself and his own advantage in the highest degree. It is undoubtedly therefore necessary that in freely aspiring to civil society he has his eye on some advantage coming to himself from it.[35]

He continues,

> Therefore the true and principle cause why heads of households abandoned their natural liberty and had recourse to the constitution of states was to build protection around themselves against the evils that threaten man from man….Respect for that law [of nature] cannot guarantee a life in natural liberty with fair security.[36]

Although natural law should regulate the behavior of human beings in their natural state of liberty, it cannot be relied on to do so. Part of the reason that it cannot properly constrain our action, Pufendorf holds, is that natural law cannot truly be enforced. As we saw above, in the state of nature no one has the appropriate authority to enforce the law of nature except God. Unfortunately, any punishments that God hands down come after death, and therefore are not effective at deterring humans from violating the law. Likewise, there are great obstacles to trusting others given that no one has the ability to punish those who break their word and betray others' trust.

For pragmatic reasons – the advantages of peace and security from harm – we must have a sovereign capable of enforcing the law and compelling individuals to follow through on their agreements and contracts. The method of this enforcement is punishment and the threat thereof.[37]

To summarize, then, Pufendorf holds that people agree to leave the state of nature and enter civil society for the purpose of ensuring the security of their persons

---

[35] Ibid., p. 132
[36] Ibid., p. 133-134
[37] Ibid., p. 140

and property. The method by which the state accomplishes this end is through the threats of punishment and the actual performance of such punishment when the necessary conditions are met. This punishment – violence committed by the sovereign against the citizens – is justified on the grounds that the original contractors would consent to it as a necessary means of achieving the state's purpose.[38]

We can look to other passages in which Pufendorf confirms that punishment is justified as means of ensuring the benefits of civil society. "Just as penalties should not be imposed except in the public interest, so the public interest should govern the extent of the penalties. In this way the citizens' sufferings will not outweigh the state's gain."[39] He also states that when punishing, the state must consider "not only what evil was done, but also what good may come from its punishment…. The real aim of punishment by human beings is the prevention of attacks and injuries."[40] Punishment can satisfy this aim in multiple ways. It can change the individual criminal by making her or him less likely to commit future crimes (either through reformation of the criminal's character or – as is more likely what Pufendorf intended – by showing the criminal that she or he cannot get away with such wrongdoing); it can effectively deter other potential criminals from engaging in illegal activity, for fear of suffering a similar fate; and it can protect the general welfare by removing (either temporarily or, in the case of capital punishment, permanently) a threat from the populace. All three of these goods can be

[38] Ibid.
[39] Ibid., p. 152
[40] Ibid., p. 159

achieved simultaneously, and all three are specifically deterrent motivations for punishing criminal activity.

Are there any indications that Pufendorf might have included some retributive elements in his theory of punishment? He does include as a criterion of punishment the following: "A Punishment is an evil one suffers, inflicted in return for an evil one has done; in other words, some painful evil imposed by an authority as a means of coercion in view of a past offense."[41] This passage has a clearly retributive character, but before we conclude that Pufendorf's position is an incoherent mixture of different justifications, let us pause to examine what exactly his statements entail. He says that punishment is "inflicted in return for an evil one has done." The best way to make sense of this is not as a justification for the use of force, but as a specification of in what situations or against whom the use of force is appropriate. By interpreting this passage as a statement about who is liable for punishment, we are capable of accommodating the backward-looking components of Pufendorf's view without contradicting the largely forward-looking nature of his position and the arguments he gives in support of it.

Pufendorf, then, seems to offer the following theory of punishment. Coercive force is justified on the part of the state on deterrent grounds. The appropriate target of punishment – who is liable – is fixed retributively, while the amount of punishment seems to depend on deterrent considerations. Pufendorf has nothing to say on the subject of the method of punishment, but given natural law theorists' pervasive use of

[41] Ibid., p. 158

"proportionality" to mean eye-for-an-eye style *lex talionis*, it would not be surprising if the method of punishment should emulate the crime committed, up to a point. If this is correct, it would imply a retributive means of fixing the method of punishment. Despite being grounded on a purely contractarian foundation, Pufendorf's theory of punishment ends up resembling a popular option from the natural law tradition discussed above: a mixed theory with a deterrent justification.

Hobbes

Roughly contemporaneous with Pufendorf's work, Thomas Hobbes published his masterpiece in social contract political theory, *Leviathan*.[42] In many ways, Hobbes represents the clearest early modern example of the strict contractarian position, although the position he stakes out on the subject of punishment is decidedly unique. His pure focus on rational self-interest, coupled with his rejection of any kind of normativity outside of the state, means that he does not recognize the need to justify the sovereign's use of coercive force in the way that other political philosophers do.

Despite his prevalence within the natural law tradition, Pufendorf still advocated a version of the strict contractarian view. This comes from his unorthodox view that natural law, though existent and normative, does not confer to humans any right to punish. Hobbes, on the other hand, starts out by taking the very concept of a natural law much less literally than his peers. While he uses some of the language, he is not

[42] Hobbes, Thomas. *Leviathan*. Ed. JCA Gaskin. Oxford: Oxford University Press, 1996.

committed to many of the core positions that are customarily associated with the natural law tradition.

In *Leviathan*, the concept of a law of nature is dramatically different from the way it is used by thinkers like Grotius, Locke, and even Pufendorf. While these other natural law theorists hold such laws to be a normative, Hobbes takes laws of nature to be little more than the products of human, prudential reasoning.[43] Given that we all have certain basic needs, he argues, and given the state of nature's rather unpleasant character, it is merely a matter of empirical fact that all humans would prefer to live in a state. In order for that to happen, certain conditions must be met. Reason, conceived of as a tool for means-ends calculations, tells us what these conditions are and how to meet them. This last step – how to meet the necessary conditions for escape from the state of nature – are Hobbes's laws of nature. They are purely a product of the human mind.

Given these commitments, it is not difficult to see why Hobbes views the purpose of the state to be determined merely as a matter of rational self-interest. We come to the state as a means of satisfying personal needs to security. The only reasons why the contractors agree to give up some of their liberty and create a state are self-

[43] The claim that Hobbes defends a completely amoral state of nature has been challenged in recent years by several significant interpreters (See Martinich, A.P. *Hobbes*. New York: Routledge, 2005.). Martinich argues that Hobbes's egoism is neither a true 'ethical' egoism nor inconsistent with the possibility of natural morality. There is not the space to address these arguments here, but by way of brief refutation I think that Martinich misses the most significant reason for holding Hobbes to posit an amoral state of nature: Hobbes's own claim that "To this war of every man against every man, this also is consequent: nothing can be unjust. The notions of right and wrong, justice and injustice, have there no place. Where is no common power, there is no law; where no law, no injustice" (Hobbes, p. 78). Egoism is not the reason for doubting Hobbes's commitment to natural morality; rather, such doubt comes directly from Hobbes's own claims that all morality and justice are only possible in social conditions with laws and an authority with the power to enforce them.

interested ones. While one could dispute the empirical claims that Hobbes makes about the nature of human motivation or what specific policies the contractors would agree to in order to achieve security, it is clear that he envisions the state as containing nothing that does not arise from the actual process of the social contract.

When it comes to the subject of punishment, however, Hobbes's position is not dramatically different from Pufendorf's. He also holds that punishment is only possible in a state with a recognized sovereign;[44]outside of the state, any attempt to punish is merely violence. Given his weaker stance on natural law, there is nothing morally wrong or unjust with such violence;[45] it merely cannot, as a matter of definition, be construed as punishment.

Punishment, then, requires a sovereign, which in turn requires a state. On what grounds is the sovereign justified in the use of violence against the citizens of the state? Again, Hobbes's answer closely resembles the one provided by Pufendorf: the contractors gave their consent to such acts of punishment in the hopes that this would secure a greater degree of safety of person and property than they had in the state of nature. The purpose of the state is to remedy the "solitary, poor, nasty, brutish, and short"[46] character of human life in the state of nature.

Although Hobbes does not use all of the terms and language that have become familiar in our analysis of early modern theories of punishment, it is the case that he employs a version of the standard deterrent justification. His stance is quite similar to

[44] Ibid., p. 85
[45] Ibid.
[46] Ibid., p. 84

Pufendorf's: the sovereign is authorized to punish on the grounds that the contractors gave him this right in order to limit – and if possible, prevent – the kind of agreement violations that become crimes in a state. Although the sovereign is meant to punish only those that violate contracts, the motivation is not retributivist; the existence of punishment makes possible agreements, on account of the deterrent effect of the threat of punishment. Hobbes is clear that the contractors have no reason to care about retributive concerns, for if we are betrayed by another, there is a good chance we will be dead and unconcerned with retribution. Instead, the contractors want to prevent such defections from occurring in the first place. Although there are objections that could be made to the argument he gives, it is clear that such arguments make Hobbes's justification deterrent.

In all other aspects of his theory, though, Hobbes differs quite dramatically from everyone else in this chapter. Given his commitment to absolutism and his belief that, in addition to punishment, justice is a construction of states, Hobbes essentially holds that any and all uses of violence by the sovereign are justified. There is nothing wrong with the sovereign's use of force against those who have not done wrong. There are no limits that can conceivably be set with respect to what degree or method of punishment the sovereign can employ. Hobbes does hold that a good sovereign will not mismanage her or his state; after all, to do so would be to weaken her or his own position, thus opening the door to attack from other states. This, however, is merely a descriptive claim, not a normative one.

Given that Hobbes essentially takes any and all violence by the sovereign to be legitimate acts of punishing, it might seem that his theory does not employ any specific justification. This conclusion, however, fails to account for the origin of the sovereign's authorization to use force. The contractors do not hand over such authority without reason, nor do they hand it over out of a desire to see wrong-doers punished. Rather, the contractors give this limitless power to punish to the sovereign in the interest of security and the prevention of crime (construed here primarily as refusing to honor one's side of a bargain). Even if the sovereign is not confined to using force in a deterrent manner, his or her authorization to use it in the first place arises for deterrent reasons.

Beccaria

This tradition of strict contractarian reasoning on punishment reaches perhaps its most compelling statement in the work of Cesare Beccaria. He does not employ much of the language of the natural law tradition, as Pufendorf did, and therefore does not need to determine whether or not the existence of natural law could give rise to punishment in the state of nature. Unlike Hobbes, he does not embrace an absolute sovereign, and as such favors a much more constrained use of coercive force. His short, focused book *On Crimes and Punishments*, written with the aim of encouraging the Austrian Lombard rulers of Milan to reform their penal system, lays out a clear case for a deterrent theory of punishment and the limitations that go along therewith. His account is progressive, unabashedly consequentialist, empirically-minded, and, like the ones discussed above,

ultimately rooted in and driven by a specific conception of what the state is and what purposes it fulfills.

In describing the state, Beccaria uses language quite similar to the other social contract thinkers of his time.

> Wearied by living in an unending state of war and by a freedom rendered useless by the uncertainty of retaining it, they sacrifice a part of that freedom in order to enjoy what remains in security and calm. The sum of these portions of freedom sacrificed to the good of all makes up the sovereignty of the nation, and the sovereign is the legitimate repository and administrator of these freedoms.[47]

The sovereign's right to punish, then, is authorized as the necessary means to protect this repository of freedoms, for in surrendering their freedom, the contractors would have seen fit to give the sovereign the power to protect it. Any punishment that is not aimed at protecting the repository of freedoms is clearly an unjustified use of force. Beccaria goes beyond this, however, to claim that any punishment that is not necessary for or efficient at producing the protection of this repository is also entirely unjustified on the grounds that the citizens would never have agreed to give up such freedoms to the sovereign when making decisions about the social contract. The citizens, as contractors, would only have consented to the state's use of the most effective and least restrictive means of ensuring the protection of the ceded liberty. In order for a punishment to be legitimate, then, it must satisfy conditions of appropriate end and of empirically demonstrable efficiency.

---

[47] Beccaria, Cesare. *On Crimes and Punishments and Other Writings*. Ed. Richard Bellamy. Cambridge: Cambridge University Press, 1995. p. 9.

What the contractors-turned-citizens have agreed to, Beccaria holds, is not that the sovereign has the right to use force against members of the state merely in light of an effort to violate the law (in Beccaria's language, to steal from the repository of freedoms). They have contracted instead that the sovereign is to protect this repository, full-stop. If we are to punish someone for having violated the repository, it is only on the grounds that doing so is necessary to protect it from future incursions (or, where possible, to restore what has been taken, or both). Punishment, then, is justified solely by the claim that it will have the effect of deterring any future crime.

Beccaria writes,

> The purpose [of punishment], therefore, is nothing other than to prevent the offender from doing fresh harm to his fellows and to deter others from doing likewise. Therefore, punishments and the means adopted for inflicting them should, consistent with proportionality, be so selected as to make the most efficacious and lasting impression on the minds of men with the least torment to the body of the condemned.[48]

While Beccaria mentions the need for punishment to be consistent with proportionality, the way in which he conceives of proportionality is quite different from a literal interpretation of equivalence between crime and punishment. He does not, for instance, endorse the idea that a punishment ought to be roughly equivalent to the crime committed in a vague, eye-for-an-eye sense. Instead, Beccaria argued that the appropriate amount of punishment was simply that which was required to outweigh whatever good was gained from the commission of the crime. "If a punishment is to serve its purpose, it is enough that the harm of punishment should outweigh the good

[48] Ibid., p. 31

which the criminal can derive from the crime…Anything more than this is superfluous and, therefore, tyrannous."[49]

In this way, he captures some of the same desire for literal proportionality that other, more eye-for-an-eye thinkers might advocate, yet there are some important differences that carry noteworthy implications. For Beccaria, the most severely punished crimes will not necessarily be those that are most terrible in the traditional sense of causing the most damage or harm. Instead, the crimes that will need the greatest penalties are those that confer the greatest good on the criminal and are therefore the most tempting. Of course, such an account has its own issues; for instance, how reliably must a punishment be able to deter criminals? How do we determine the relative attractiveness or temptingness of certain crimes?

While these are clearly issues that a state implementing Beccaria's view of punishment would need to address, they are not insurmountable problems. Take, for instance, the problem of how much punishment is necessary to deter crime. We might ask: are criminal behaviors not already punished in Beccaria's time? And yet, despite penalties, people engage in such criminal activity. Clearly, the punishments are not suitably effective at deterrence. Should we continue to make the punishments harsher, on the grounds that not everyone is deterred? But if this is the solution, then is it possible that we run the risk of going beyond what is justified, since the punishment is harsher than what is necessary to deter the average person? Beccaria's response to these

[49] Ibid., p. 64

set of problems focuses on modifying punishment, but not by making it harsher. Instead, he argues, we need only to increase the certainty and regularity of punishment. He writes, "One of the most effective brakes on crime is not the harshness of its punishment, but the unerringness of punishment….The certainty of even a mild punishment will make a bigger impression than the fear of a more awful one which is united to a hope of not being punished at all."[50]

In addition to revising the justification and amount of punishment in a more deterrent and, he hoped, humane direction, Beccaria also tackled the subject of what kinds of punishments were permissible for state use. In other words, unlike Pufendorf and Hobbes, he does comment directly and extensively on the method of punishment, even if his comments are merely the imposition of a few limitations on what methods of punishment the state can employ. While most natural law thinkers had not commented on this issue beyond simple calls for proportionality between crime and punishment, Beccaria devoted significant energy to arguing against the use of capital punishment, torture, and excessively harsh physical or corporal punishments. As we will see, each of these arguments ultimately derives its force from his use of deterrence as the general justification for the state's use of coercive force against its citizens.

On the subject of capital punishment, Beccaria gives several arguments. The first argument is one that goes back to the heart of Beccaria's strict contractarianism: no contractor would agree to give up the power to end her or his life to the sovereign. In

[50] Ibid., p. 63

one sense, he seems to be claiming the right to life to be inalienable: no one is capable of waiving such a fundamental right, meaning that the state cannot claim to have such a power. Not content with this claim, however, he also offers support in the form of a kind of rational choice theory argument: people enter civil society with the aim of collecting security of person and possessions, and no one would be foolish enough to give up his or her life in order to obtain such security.[51]

Based on this argument, Beccaria concludes that if the state executes an individual, it must be an act of war.[52] The citizen must commit a crime of such a nature as to remove herself or himself from the state, or else one that threatens the very life of the state by its commission. In such a case, the state no longer executes a citizen, but instead engages in an act of war against an external enemy. This solution is not available in most cases, however; once again, we can look to the contractors to see that they would not consent to most crimes being ones that resulted in a loss of citizenship, thereby making execution a permissible crime.

He expresses a similar worry when arguing against the use of harsh physical punishments. He has in mind here particularly back-breaking labor or punishments that involve mutilating the criminal in some way. Not only does the use of especially harsh penalties for violating the law cause a society to itself become more violent,[53] but they are not even effective. Recall the quotation provided above, concerning the relative effectiveness of harsh punishments versus extremely regular and reliable punishments.

---

[51] Ibid., p. 66
[52] Ibid.
[53] Ibid., p. 72

Put simply, risk takers will still violate the law if they think there is a chance they will not get caught. Further, the institution of particularly harsh punishments will cause criminals to commit further crimes in their desperation to avoid capture and punishment. All harsh punishments do is increase suffering and violence without any demonstrable benefit, an outcome that is unacceptable to Beccaria's consequentialist outlook.

Finally, he argues against the use of torture on similarly pragmatic grounds. The practice of torturing criminals, he holds, fails to achieve any desirable end. Not only does it fail on the same grounds as harsh physical punishments, but it also fails to achieve the special goal that was frequently given as a justification: the acquisition of information. When torture is used to gain confessions, admissions of involvement in past crimes, or accusations against other wrongdoers, Beccaria argues, the information received is highly unreliable. Those being tortured will say anything to make the pain stop.

Of all the thinkers examined thus far, Beccaria is the clearest and most self-conscious example of deterrence. While many make retributive sounding claims despite their justifications, which rely on deterrent arguments, Beccaria approaches the subject of punishment with the goal of scientific precision. There is no difficulty reconciling various aspects of his position, and there is a great deal of internal consistency in the way in which he approaches practical methods of punishment. He demonstrates that as long as one is willing to accept the premises of the strict contractarian position, there is a perfectly workable way of constructing a theory of punishment available.

2.3 Punishment as a State Construction: The Normativists

The strict contractarians, however, are not the only tradition of thinking about punishment as a construction of states. There is a second sub-group of philosophers that also holds punishment to be possible only in the context of laws and executive authority, but maintains that we can know at least some of the features the state must adopt independently of considerations of what contractors would agree to. In some cases, the state's purpose is set by human teleology: humans have a specific end, the state exists as a necessary tool to help humans achieve this end, therefore the state must be designed in such a way as to guarantee its effectiveness at aiding in the attainment of humankind's end. In other cases, human beings have a moral duty to enter the state, as such a set of institutions are the only way to guarantee moral or rightful conditions, and therefore any state must have the right kind of makeup to ensure the creation, promotion, and protection of these conditions. What both of these kinds of views share is their basic commitment to the idea that there are normative reasons for the state to exist and for humans to enter and remain in it. These normative reasons, in turn, tell us what kind of state we have a duty to create.

Rousseau

One of Kant's most significant influences in political theory,[54] Jean-Jacques Rousseau developed an account of the state and political legitimacy that emphasized the role that hypothetical, rational consent plays in determining the . While earlier social contract thinkers like Hobbes, Pufendorf, and Locke had never truly thought of the social contract as a document to be signed, Rousseau introduced in his *On the Social Contract* an additional degree of abstraction, according to which the social contract becomes more explicitly a deliberative standpoint to be occupied by political agents.[55] His approach – and, in particular, the concept of the 'general will' – not only contributed to Kant's own views in political philosophy, but it has also continued to inspire theorists through to today.

Given the shorter, more focused nature of *On the Social Contract* and his primary focus on combatting inequality, Rousseau's views on punishment are less fully and carefully detailed than are some of the other views we have explored. Nevertheless, there is enough material to situate him within the framework that I have been developing. While his theory faces some internal difficulties, its broad contours are similar enough to the ones we have previously explored: a strong deterrent justification underlies and supports his arguments for the permissibility of the state's use of punishment.

---

[54] Kant was famously so enthralled by Rousseau's *Emile* that he missed his customary afternoon walk. For a fuller account of Rousseau's influence on Kant, see Cassirer, Ernst. *Kant's Life and Thought*. James Haden, trans. New Haven: Yale University Press, 1981, pp. 86-90.

[55] Rousseau, Jean-Jacques. *On the Social Contract*. Donald A. Cress, trans. *The Basic Political Writings*. Indianapolis: Hackett Publishing Company, 1987.

For Rousseau, the state of nature is not characterized by morally obligatory natural laws, promulgated by an authoritative God. He does suggest that justice flows from God, but that we are incapable of understanding it without the mediation of states and laws.[56] Without them, he believes, like Hobbes, that a person in the state of nature has "a right to everything that tempts him and he can reach," and that it is only after entering civil society that justice becomes a reality.[57] This is not to say that Rousseau believed that the state of nature was a fully amoral condition, like Hobbes. Rousseau's version of natural normativity might not include the concepts of duty-laden and law-like constraints, but he does outline a perfectionistic ethics that direct human behavior toward the development of certain characteristic capacities and talents.[58] Rousseau is also relatively unique for his time in holding that these capacities are largely socially nurtured. This is the reason I have classified Rousseau as a 'normativist;' his belief that only certain kinds of societies can provide their citizens with the necessary social conditions to achieve (or, at least, approach) the perfection that he holds to be the highest good of human life indicates his commitment to the view that the appropriate form and purpose of the state is guided by more than just the rational self-interest of the contracting agents.

Given that Rousseau's conception of morality in the state of nature does not include 'laws' and strict obligation, it is no great surprise that he does not conceive of

[56] Ibid., p. 160
[57] Ibid., pp. 150-151
[58] Despite similarities in their political theories, Kant and Rousseau share little in the way of positions in moral philosophy. While Kant also believes that people should perfect their capacities and talents, he construes this as a duty, and indeed, as one that is secondary to other, more law-like duties.

punishment as being a natural phenomenon. Punishment might arise between individuals living in an inegalitarian, non-contractual society – of the sort that Rousseau discusses in the *Discourse on the Origin of Inequality*[59] – but this kind of punishment would be inherently unjustified. Those punishing would be truly engaging in an act of self-defense or war, not punishment. As Rousseau states when discussing the supposed 'right of the strongest,' the power to use force does not confer upon anyone a moral permission to do so.[60] Given his concerns about the dangers that result from the loss of one's independence, even submitting to the will of another for rational, prudential reasons does not confer upon them the right to use force to punish.

It is only be the specific, reciprocal process by which the social contract is created that individual people are capable of entering into the kind of social arrangement wherein real punishment becomes possible.[61] Rousseau keenly points out that if the social contract requires that a people must give consent to political authority, this implies that the people exist, as a community, prior to the institution of political authority.[62] The social contract, then, is not the establishment of a ruler, form of government, or even political constitution; rather, the social contract is the agreement by which disparate individuals become a political community together. Only such a community has the power to create a constitution or government, and it is only through such self-legislation that coercive authority can be exercised. When it comes to the extent

---

[59] Rousseau, Jean-Jacques. *Discourse on the Origin of Inequality*. Donald A. Cress, trans. *The Basic Political Writings*. Indianapolis: Hackett Publishing Company, 1987.
[60] Rousseau, *Social Contract*, p. 143
[61] Ibid., pp. 147-150
[62] Ibid., p. 147

of the permissible use of coercive force, Rousseau's answer closely follows Hobbes's. Rousseau argues that only the "total alienation of each associate, together with all of his rights, to the entire community"[63] can ensure that independent individuals can come together as a unified body politic. What this means, however, is that there are no strong checks – no inalienable rights – to keep the actions of the sovereign in check. True, the sovereign is governed by the general will, which aims unfailingly at "public utility."[64] All this means, however, is that the state cannot adopt policies that cause useless harm. Besides this, the state has "an absolute power over all its members"[65] It can, and indeed must, sacrifice the minority for the majority when doing so promotes the general welfare.

It is for this reason that Rousseau's justification for punishment is best characterized as deterrent. The state punishes so as to decrease the number of future crimes against the general welfare. The sanctions imposed on any particular criminal are merely the acceptable costs of achieving this result. If a person cannot be rehabilitated, they can at least be made an example of, so as to decrease the likelihood of others engaging in wrongdoing.[66] It is also worth noting that Rousseau includes an interesting statement about the appropriate amount or limitations of punishment. While he believes that the death penalty is not inherently problematic, he does suggest that it can only be used when the criminal "cannot be preserved without danger."[67] This indicates that

[63] Ibid., p. 148
[64] Ibid., p. 155
[65] Ibid., p. 156
[66] Ibid., p., 160
[67] Ibid.

while deterrence is the central justification for Rousseauian punishment, it is also constrained by strong rehabilitative elements as well.

Rousseau does have a small amount of additional material on punishment; unfortunately, it introduces some significant problems for his account. In several key passages, he indicates that the commission of any crime, regardless of how small, is a violation of the social contract and grounds for the expulsion of the perpetrator. After this point, the state responds to the banished member as it would a foreign enemy who seeks to make war on the commonwealth. In effect, this position ultimately entails the elimination of the entire institution of punishment; all that remains is the state's power to wage war.

This power is manifested in the way in which Rousseau envisions "punishment" as being carried out in civil society. He writes,

> Every malefactor who attacks the social right becomes through his transgressions a rebel and a traitor to the homeland; in violating its laws, he ceases to be a member, and he even wages war with it. In that case, the preservation of the state is incompatible with his own. Thus one of the two must perish; and when the guilty party is put to death, it is less as a citizen than as an enemy. The legal proceeding and the judgment are the proofs and the declaration that he has broken the social treaty, and consequently that he is no longer a member of the state.[68]

Although he uses the term punishment elsewhere in *On the Social Contract*, in this passage Rousseau reveals that he does not leave any conceptual room for any institution of punishment. Clearly, finding a citizen guilty of crime is sufficient for revoking that individual's citizenship.[69] If punishment is understood as the use of coercive force

---

[68] Ibid., p. 159

[69] There is some ambiguity as to what Rousseau takes to be the *cause* of the loss of membership in the body politic. One possibility, which we can call the 'revocation reading,' is that the perpetrator of

against citizens in response to some violation of positive law, then it is clear that no such thing can occur in Rousseau's vision for the state. One cannot be both a citizen and an appropriate subject of punishment; while one is a citizen, punishment would be inappropriate and undeserved, and once punishment is appropriate and deserved, the target is no longer a citizen. Instead, the violence visited upon the criminal is not conceptually different from the force the state would use against an enemy.

Rousseau's theory of punishment is undermined by his premise that every violation of law is also a violation of the social contract. While he is not alone in making this kind of error, it is not an issue that Kant faces in his own political writings. The idea that every violation of positive law is sufficient to revoke one's membership in and obligation to a state or community is not among the many Rousseauian elements that he incorporated into his views. Nevertheless, Rousseau is still yet another of Kant's primary influences in practical philosophy who supported a deterrent justification for punishment.

Smith

Aside from Kant, the eighteenth century philosopher most frequently associated with the retributive school of thought is Adam Smith. This is not without good reason:

---

crime has her or his citizenship revoked when she or he is found guilty of committing a criminal act. In this case, the loss of membership occurs as an official function. The other possibility, which we can call the 'recognition reading,' is that the criminal loses her or his membership as soon as the crime is committed. In this case, the court does not revoke a criminal's citizenship; it merely recognizes that it has already been lost.

Smith's theory of moral sentiments clearly establishes that violence is morally justified as punishment when it is met with or motivated by the appropriate retributive feelings from others adopting the perspective of the impartial spectator. This, however, is not all that he has to say on the subject; indeed, he also introduces another class of punishments that do not meet the criterion of an appropriately sharable sentiment of resentment. The sovereign can, according to Smith, make blamable and punishable those offenses that do not, in and of themselves, generate the requisite kind of affective state in the individual or others adopting the perspective of the impartial spectator. The sovereign has this power as it is necessary for the stable functioning of the state – essentially, for the sake of utility. I will argue, following Knud Haakonssen, that the tension between the retributive and deterrent elements in Smith's theory of punishment are resolvable by demarcating the categories of individual, moral behavior and the just actions of states.

Briefly, the traditional account of Smith's theory of punishment is focused on the sentiment of resentment. As opposed to other negative, hostile sentiments – like hatred, for instance – resentment is what we feel toward one who has done us wrong.[70] Smith describes actions that cause us to feel resentment as having the quality of demerit – that is, of deserving punishment. Violent acts that are not motivated by resentment cannot be truly or accurately described as punishment.

Simply feeling resentment is not enough; the resentment must also be appropriate. In order for resentment to be morally appropriate, it must pass the same

[70] Smith, Adam. *The Theory of Moral Sentiments*. Indianapolis: Liberty Fund, 1982. p. 69

test as all of Smith's other sentiments: namely, it must be one that can be shared by another agent who adopts the perspective of the 'impartial spectator.'[71] While fully fleshing out all the details of the impartial spectator would take more space than this chapter can allow, for our purposes it is enough to understand that Smith describes the impartial spectator as a hypothetical perspective we take on when considering how we would view an action or situation that did not personally affect us in any significant way. Once abstracted away from our personal commitments, we would all have similar affective responses to the same kinds of behaviors or situations. To bring this back to resentment, Smith would argue that my feelings of resentment toward a person who attacked me without provocation would be the kind of resentment that anyone viewing the situation from the perspective of the impartial spectator could share. If I resented someone for accidentally and faultlessly stepping on my toe as I rushed carelessly through the street, then this would not be a sharable sentiment. The person in question would not deserve to be punished, unlike my attacker from the last example.

It is important to note before moving on that the above justifications for punishment do not take into account the utility that the punishment generates for society. While Smith does claim that the system of retributive punishing based on sentiments of resentment leads to a highly beneficial state for human beings,[72] he explicitly rejects the idea that it is utility which justifies the punishment in the first place. He writes,

---

[71] Ibid., pp. 17-19
[72] Ibid., p. 86

> And with regard, at least, to this most dreadful of crimes [murder], Nature, antecedent to all reflections upon the utility of punishment, has in this manner stamped upon the human heart, in the strongest and most indelible characters, an immediate and instinctive approbation of the sacred and necessary law of retaliation.[73]

Here, Smith is claiming that our resentment is anterior to any considerations of utility. Presumably, then, an individual would still experience resentment, the impartial spectator would still approve of it, and punishment would still be justified even if resentment toward and punishment of a given wrong-doing were in opposition to the greatest possible utility. Smith has already established that our passions can run counter to utility.[74]

In addition to this resentment-based, Smith also offers another view of punishment that seems to conflict with his prior statements. Although it is true most of the time that there can be no punishment without the passion of resentment, this does not hold toward the end of Smith's writings on punishment in *The Theory of Moral Sentiments*. In the clearest example of this alternative approach, he claims "When a sovereign commands what is merely indifferent, and what, antecedent to his orders, might have been omitted without any blame, it becomes not only blamable but punishable to disobey him."[75]

This complicates the picture that Smith has been depicting. Up to this point, Smith's view of violence held it to be only justified as punishment when it met the

[73] Ibid., p. 71
[74] Ibid., p. 35
[75] Ibid., p. 81

affective standards that I have described above. Now, he seems to be claiming that the sovereign has the ability to declare an action deserving of punishment, regardless of whether or not it previously elicited such feelings. He offers an example of a situation in which the sovereign renders an action punishable: a soldier who falls asleep while on watch duty. The soldier's inability to stay awake likely causes no direct harm, and no other individual can claim to be wronged by it. To put this into Smith's language: there is no one who could claim to have the right kind of retributive affective response (i.e. one that others adopting the perspective of the impartial spectator would share) to the solder's having fallen asleep on duty. We might feel some kind of disapprobation for the soldier's actions, but not the kind that Smith has previously identified as being the basis for justifiable punishment. Nevertheless, he holds the sovereign as able to punish such an action.

What is at work in this separate class of cases? Smith offers us a clue when he claims that the traditional punishments founded on retributive sentiment and these special cases are "far from being founded upon the same principles."[76] Although he does not state directly what the principle underlying these politically necessary cases of punishment, I submit that Haakonssen's account is essentially correct: the principle of utility justifies the punitive actions of states, as opposed to the sentimental, moral principles that guide the behavior of individuals.[77] As such, there are actually two separate kinds of actions being described. Despite the fact that we call them both

[76] Ibid., p. 91
[77] Haakonssen, Knud. *The Science of a Legislator*. Cambridge: Cambridge University Press, 1981. Pp 114-123.

punishment, the legal form of punishment bears no relation to and is not founded upon the punishment that is conceivably possible in the 'state of nature.'

First, we saw above that it is an essential feature of individual punishing that it be motivated by resentment. Furthermore, part of what makes resentment a discreet sentiment, distinct from hatred, is that when I experience resentment toward another, I want to be the cause of this other's suffering. While my hatred can be satisfied if he or she merely meets with an unhappy accident, this does nothing to placate my resentment. Resentment can only be satisfied if I bring about his or her suffering. Similarly, if I share in another's resentment, then I want her to be the cause of her target's suffering.

This, however, is decidedly not how punishment occurs in a state. Indeed, the state exists in part to prevent this personal method of delivering suffering to others who have done us wrong. Instead, the sovereign is now directly and solely engaged in the process of punishment. All victims, regardless of how sympathetic their resentment is, must watch the state be the cause of the wrongdoers' suffering. If state punishment were founded on the same principle that governs individual punishment, this would not be possible. There would be no state punishment.

The second clue we have to what different principles could be motivating the state's different use of punishment is observable in the 'laws of police.' These are the kind of pedestrian policies regarding maintenance of the state that are clearly necessary for the smooth functioning of civil society, and yet are not the sort of laws that commonly strike one as concerning matters of justice. The only way to make sense of

such policies, argues Haakonssen,[78] is by conceding that the state functions on a separate system, independent of what the morality of individuals requires or allows. The state is not, contra the natural law theorists, a 'moral person'" and is incapable of experiencing sentiments; thus, it is guided exclusively by an interest in the common good or utility.

These different principles also explain why Smith does not belong with the first group I considered, the natural rights theorists. Recall that this group holds 1) punishment is possible for individuals in the state of nature and 2) the justification for the state's use of punishment is the same as the individuals' in the state of nature. Although Smith seems to suggest that 1) is the case, he does not endorse 2). Punishment in the state of nature is justified by reference to the appropriate kind of moral sentiments. Punishment in civil society, however, is justified by its necessity in achieving the utilitarian ends of the state. As such, the kind of punishment that the state employs is itself a construction of state authority; individuals cannot legitimately punish on the same kind of basis that the sovereign uses.

This is true of not only the definition, but the justification as well. Given the affective requirements of appropriate individual, pre-civil punishing, we might conclude that such cases are indeed justified in a retributive manner. The interest of the state, however, is clearly a deterrent one. This aligns with the sentiments the impartial spectator has with respect to the sleeping sentry: we do not blame or think the sentry is deserving of punishment, but we recognize that the state has an interest in preventing

[78] Ibid., p. 116.

such behavior. Insofar as this dissertation is focused on the institution of punishment as it occurs in states, it is this latter form of punishment that matters most for our purposes. And in this respect, Smith is clearly among those who hold punishment to be justified by deterrent concerns.

Conclusion

Throughout this chapter, I have made the case that political philosophy of the 17th and 18th centuries advanced a number of defenses of the state's use of coercion and violence against its citizens. Broadly, I have divided these different strategies of grounding the right to punish into two main categories, distinguished by one feature of the definition of punishment they offer: namely, whether or not punishment requires a state. Despite the differences present in these two major positions, they are both united by a common theme: they employ a specific conception of the state that in turn requires their use of some form of deterrent justification of the institution of punishment. This deterrence is then buttressed by a variety other retributive and/or rehabilitative constraints and interests.

The dominance of this deterrent form of justifying punishment was a relatively recent development. While ancient Greek philosophy had been characterized by a number of rehabilitative justifications and the theologically-driven medieval period had favored retributive theories, the early modern period saw the rapid emergence and development of deterrence. This shift was largely inspired by new ways of thinking about the role of government and the separation of political and moral spheres. In the

face of rampant religious and sectarian wars, philosophers and jurists sought forms of political organization and justification that could place the state on stable, peaceful foundations. In the search for a version of civil society that could be reasonably accepted by all, these thinkers turned to states that refrained from explicit moral evaluations of its citizens. This new paradigm virtually entailed a shift toward preventive, efficiency-oriented punishment.

Where does Kant fit into this picture? I have claimed that he belongs to the normativist tradition within the school of thought that holds punishment to be a construction of states. I will aim to prove this claim in the following chapter, but if it is granted for the time being, then we can already see that Kant fits roughly within the broad trends of his time. The originality of Kant's practical philosophy exists within this framework; at no point does he attempt to develop a wholly new or radical approach to understanding, forming, or justifying the state and its relationship with its members.

In fitting in with the basic contours of his predecessors, Kant takes on more substance than one might think. The move toward secular, liberal states that unites the natural law theorists, the social contractarians, and the early consequentialists includes a shared assumption that states are not in the business of evaluating and responding to the moral character of citizens. Rather, states exist in order to preserve security, promote the happiness of their members, or establish conditions under which the members can develop as moral, progressive beings. In all cases, these aims are furthered by the effective prevention of instances of crime. The view that the state would have an interest in or even an authorization to respond to moral desert is both alien to this tradition and

highly difficult to accommodate. While Kant will want to make such an argument himself, it seems as though he has underestimated the degree to which the basic elements of political philosophy that he has adopted from his predecessors have closed the door on such a possibility.

3 Defining Punishment: Coercion and Right

The first task for any theory of punishment is to establish a definition of punishment. This will enable the theory to pick out which acts of violence or coercion count as punishment and which do not. While the justification offered for punishment will always be the most significant aspect of any theory, it is still crucial that we have a firm understanding of what kinds of actions, practices, policies, and institutions the theory is aiming to justify. In order to achieve a greater understanding of the philosophical issues pertaining to punishment, we must first have a clear picture of the phenomenon in question.

As we saw in the previous chapter, Kant's predecessors and influences defined punishment in a wide range of different manners. Some saw punishment as a coercive response to the violation of natural law, while others viewed it as a state-constructed institution that could not exist outside of civil society. Although Kant is occasionally identified as a natural law theorist,[1] at least in this respect he differs from the standard natural law view; according to Kant, punishment is neither possible in the state of nature, nor based directly upon a power or right held by individuals in the state of nature. Rather, he consistently uses punishment to refer exclusively to an executive power of states that is only possible under specified civil conditions.[2]

[1] See Mulholland, Leslie. *Kant's System of Rights*. New York: Columbia University Press, 1990. Pp. 10-15.

[2] I also claimed that Kant belongs to the 'normativist' camp, meaning that he holds there to be specific normative requirements that necessitate the existence and shape the structure of the state. This aspect

The goal of this chapter is to fully examine the conditions that must be met for an instance of punishment to occur. For Kant, this is a relatively high standard, as punishment can only take place within an institutional framework. I contend that Kant defines punishment as a coercive action, undertaken against a citizen of a state by the legitimate executive, as a sanction for the violation of public law.

In the first section of this chapter, I work through the various components of Kant's definition, highlighting their textual support and any technical nuances in usage. Given that Kant does not offer a specific definition for punishment, this section aims to put together the various pieces that he leaves scattered throughout his work. I provide further support for this definition in the second section of the chapter. Here, I consider the division between duties of right and duties of virtue that Kant establishes in the *Metaphysics of Morals*. As I argue, the basis for this division centers squarely on the possibility of punishment for failure to perform duties of right. Finally, in the third section of the chapter, I investigate the subject of using rewards instead of punishments to incentivize behavior. Kant rejects this prospect – which he calls allurement – on the grounds that he conceives of law as coercive. As I show, he lacks a substantial basis for rejecting the use of reward to incentivize external compliance with the law. Instead, he should accept such a possibility as justified on the same grounds as punishment.

---

of his practical philosophy, however, will not play an important role until the next chapter. For more on why Kant is a normativist, rather than the more voluntaristic 'strict contractarians,' see Kersting, Wolfgang. "Kant's Concept of the State." *Essays on Kant's Political Philosophy*. Howard Williams, ed. Chicago: University of Chicago Press, 1992, pp. 147-148.

Regardless of his inconsistency on the issue of reward, Kant's definition of punishment is clear, robust, and consistent with his underlying practical philosophy. He connects it up to his fundamental political and legal positions in a way that is both well supported by them and, in turn, mutually reinforcing of them. By focusing on punishment as a legal institution, deriving its authority from the general legislative will, Kant is able to situate punishment prominently as one of the distinctive features of the authority characteristic of liberal republics.

3.1 Kant's Definition of Punishment

At no point in his published writings on practical philosophy does Kant give a clear, direct definition of punishment. He does, however, give numerous indications of the primary criteria of the possibility of an act of punishment. Additionally, he has several more explicit statements in his lectures that are worth considering. Altogether, these elements paint a relatively clear picture of how Kant is using the term 'punishment': a coercive action, undertaken against a citizen of a state by the legitimate executive, as a sanction for the violation of public law.

This definition has several discreet elements, each of which deserves consideration in turn. To begin, then, we need an account of what Kant means by the term coercive action, or more generally, coercion. In its traditional usage, coercion is associated with one party compelling another's performance (or non-performance) of a particular action through the use of force, threats, or other non-rational means of compulsion or persuasion. Although coercion need not necessarily force the target to act

contrarily to her will, it has the capacity to do so. For instance, I might have been predisposed to perform action x prior to someone coercing me to perform action x. In such a case, I would still be coerced, and what would have been a freely chosen action prior to the coercion becomes unfree in some important respect. For this reason, coercion is typically thought to diminish the patient's freedom (with Hobbes representing a notable exception to this dictum), and as such coercive acts are *pro tanto* wrongful.[3]

Kant's view on coercion is in keeping with these traditional definitions, albeit characterized by his own specific understanding of freedom and the will. According to Kant, "Any action is *right* if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law" (6:320). This is what he calls the Universal Principle of Right (UPR), and it guarantees the right of each individual to exercise her or his external freedom in a way that does not violate any other's right to do the same. This entitlement to the use of our own external freedom is the only right that individuals hold in the state of nature; all other rights are either derived from UPR or are legal creations of states.

Kant understands this external freedom as our capacity to set and pursue ends. By 'set and pursue ends,' Kant has in mind more than simply the ability to wish that

---

[3] For more detailed accounts of coercion, see Gorr, Michael. "Toward a Theory of Coercion." *Canadian Journal of Philosophy*. Volume 16, Number 3, Sept. 1986, pp. 383-406; Nozick, Robert. "Coercion." *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. Sidney Morgenbesser, Patrick Suppes, and Morton White, eds. New York: St. Martin's Press, 1969, pp. 440–472; Pallikkathayil, Japa. "The Possibility of Choice: Three Accounts of the Problem with Coercion" *Philosophers' Imprint*, Vol. 11, No. 16 (November 2011), pp. 1-20; and Pettit, Philip. *A Theory of Freedom*. Oxford: Oxford University Press, 2001.

something were the case. In order to truly describe a person as free to do *x*, it must be within his power to. While Kant holds this external freedom to be less significant than the radical, inner freedom that we all possess as beings with a noumenal dimension, it is far simpler to successfully appraise the manner in which it is used. As we will see in the next section of this chapter, whatever role metaphysical freedom plays in Kant's moral philosophy, his political philosophy concerns itself with external freedom.

Arthur Ripstein argues that, for Kant, coercion should be understood as any action that limits freedom.[4] He defines freedom as "independence from being constrained by the choice of another person,"[5] and this is indeed the way in which Kant describes freedom at times (6:237). This definition works well in situations characterized by social interactions between persons. There are potential difficulties with this definition, however, that should be addressed. For instance, my capacity to set and pursue ends might very well be constrained by another making perfectly appropriate, permissible choices. This is a problem that Ripstein anticipates, and he has a very simple solution available: freedom is independence from being constrained by the non-rightful choice of another.[6] When my neighbor choses to remove a tree from her property, her actions are in accordance with UPR and therefore rightful. Thus, even if my will is constrained – I can no longer sit in its shade, perhaps – my freedom is not interfered with or diminished.

---

[4] Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge: Harvard University Press, 2009. P. 54.
[5] Ibid.
[6] Ibid., p. 55

Nevertheless, I think there are reasons for considering Ripstein's definition as too limiting. Given that he has defined coercion as always involving the choice of another person, he is unable to classify a wide range of potentially constraining external conditions as destructive to one's freedom. Imagine a case in which a person falls into a deep hole and is unable to escape.[7] In such a case, the agent might well not be constrained by the choice of another, yet it seems strange to describe her as free. Alternatively, we might describe a legal or bureaucratic institution as coercing a person, even if no particular person ever constrains her choice. Being independent of others' arbitrary choices might be a necessary condition for freedom, but I submit that it is not a fully sufficient way to characterize Kantian freedom. Instead, we ought to employ a more positive conception that focuses less on the arbitrary wills of others and more on the conditions for autonomy of the agent in question. This more fundamental definition would describe of freedom as the capacity to set and pursue ends, as it is this capacity with which being subject to the will of others interferes. This more basic definition also has the additional advantage of being able to accommodate the kinds of cases in which a person is coerced without having her choices subjected to the arbitrary will of another.

Coercion, then, is the action of another that constrains one's external freedom by interfering with her ability to set and pursue her own ends. Presented in this light, it might appear that all instances of coercive action are wrong, as they violate another's autonomy and most fundamental right. We should avoid this conclusion, though, as the

[7] This example is inspired by Raz's "man in a pit" case. See Raz, Joseph. *The Morality of Freedom*. Oxford: Clarendon Press, 1986.

right we have to the exercise of our external freedom is only guaranteed provided that our action can coexist with everyone else's external freedom, according to some universal law. If I choose to act in some non-universalizable way (e.g., I attempt to rob another by force), then anyone who prevents my action coerces me. This coercion, however, is justified in Kant's eyes, on the grounds that coercion used to prevent coercion is perfectly consistent with the UPR. He writes,

> Resistance that counteracts the hindering of an effect promotes this effect and is consistent with it….If a certain use of freedom is itself a hindrance to freedom in accordance with universal laws (i.e., wrong), coercion that is opposed to this (as a hindering of a hindrance to freedom) is consistent with freedom in accordance with universal laws, that is, it is right. Hence there is connected with right by the principle of contradiction an authorization to coerce someone who infringes upon it. (6:231)

There are, according to Kant, many forms of justifiable coercion and violence that take place between individuals, either within a state or in a hypothetical, pre-civil condition. Instances of self-defense, for instance, are appropriate ways of coercing another's respect for one's own external freedom – or, to put it another way, it is a hindrance to the hindering of one's freedom. Among these instances of justifiable violence, there are even some that appear to be punitive. Consider two examples. First, imagine the case of a parent imposing some penalty on his child in response to the child's impermissible behavior. Second, imagine the case of one or more individuals using violence to subdue and neutralize a person who has been harming others. In both these cases, some use of coercion interferes with a person's freedom in response to a specific act. While the ends, motivations, and maxims of the 'punishing' parties might be

different, the actions seem to fall under the same general description. In light of this, we might be highly tempted to describe each case as an instance of punishment.

To call these examples instances of punishment, however, would be to deviate from the standard Kantian view. Although they are coercive, they are lacking in the other necessary attributes of punishment. For instance, in the case of both the parent and the vigilantes, the authority that is being exercised is fundamentally different from the kind of authority that a state holds over its citizens; it is this latter kind that is necessary for the existence of punishment. To fully understand the difference between a parent's authority over his or her child and the kind of authority that enables punishment, we need to examine Kant's concept of a 'rightful condition.'[8] Kant does not directly define the rightful condition, but he explains that it is the state of affairs that a functional legal system is designed to bring about (6:311). The term rightful condition can be interpreted in several ways, but throughout the dissertation I will be using it in its minimal sense, to mean the necessary circumstances for individuals to have duties of right, or legal duties, to their peers. While a perfectly just state would clearly be a rightful condition, less just states might possibly be rightful as well. There is no definite line between rightful states and those that are not, but we can nevertheless point to features that the former must have or almost always have.

[8] For a full treatment of the rightful condition in Kant's political philosophy, see Byrd, B. Sharon and Hruschka, Joachim. *Kant's Doctrine of Right: A Commentary*. Cambridge: Cambridge University Press, 2010. Pp. 24-33.

One of the key conditions of the possibility of a rightful condition, according to Kant, is the existence of a general legislative will (6:320). Although there are many possible forms of legislative arrangement within a civil society, Kant writes that "legislative authority can belong only to the united will of the people" (6:313). This does not mean that all legislative decisions must be made democratically, but rather that all laws – and the process that creates them – must be consistent with the rational wills of each and every citizen. This imposes certain constraints on the legislative arrangement and the kinds of laws it can produce; for instance, a true legislative authority lacks the power to create a law to which even one citizen could not rationally consent (8:304; 6:329-6:330). In order for a rightful condition to possibly exist, there must be public laws, and in order for such laws to exist, there must be a legislative authority that arises from the general, united will of the people.[9]

All of this indicates that any laws passed by the united legislative will of a people must have several characteristics. First, they will be general. The generality of law is one of the features that distinguish it from an edict or decree. While a despotic king can pass limited, individual judgments on particular cases, a law must apply broadly. This generality of law has two distinct senses. On the one hand, these laws will be general in the sense that they must apply to all persons or classes of persons (e.g., citizens or non-

[9] There is some evidence to suggest that any state, regardless of its legislative makeup, will come closer to the ideal of a fully general legislative will than the conditions in the state of nature. Kant seems to support such thinking in his writings on the impermissibility of revolution – a subject to which I will return in chapter seven of the dissertation. Even if we were to grant that any state is better than no state, this still does not demonstrate that any form of legislative power is sufficient to guarantee the existence of a rightful condition.

citizens) without exception. No law can be made that only applies to specific individuals. On the other hand, public laws must be general in the sense that they apply in all circumstances. The law, by definition, cannot have exceptions or instances in which it does not command with authority.

In addition to their generality, public laws must also form a consistent system. It would not be possible, in other words, for an action to be permitted by one law and forbidden by another. For such a contradiction within the content of the law to be possible, the laws would necessarily have originated from a will that was not rational or governed by the right considerations. A true legislature, then, can only create laws that are wholly consistent with one another.

Third, it is a necessary feature of law that it be publicly known. There cannot be laws that are made in secret, compelling or forbidding unaware citizens from performing certain actions. The fact that one or more citizens might be unaware of the existence of some law does not invalidate it, but its having been legislated in secret does. It is possible that secret laws could still be rightful, but they could not obligate citizens with a duty to obey.

Finally, all public laws must carry with them a specified sanction for violation of the law. While contemporary legal philosophy is hardly united in its stance on relationship between law and sanction,[10] Kant clearly expresses the view that law is analytically connected to the concept of coercion and sanction (6:219). As such, any law

[10] For an example of the view that laws without sanctions are a conceptual possibility, See Hart, HLA. *The Concept of Law*. Oxford: Oxford University Press, 1962. Pp. 20-25.

passed by the legislative authority must specify the sanction to be imposed for its violation. A state that had only laws and no sanctions would, according to Kant, be no state at all.

Having a general legislative will and the laws that are produced by it, however, is not enough alone to guarantee a rightful condition and therefore the possibility of just punishment. Even if the laws all contain sanctions, these sanctions are meaningless without an executive power with the authority to enforce them. Although Kant's republican separation of powers clearly holds the executive to be subordinate to the legislative (6:313), he holds that there must be a separate executive figure to enforce the law (6:317). If the legislative sovereign also took responsibility for directly punishing, the governing authority would have descended into tyranny.[11] While Kant is very clear that such a government must still be obeyed by its citizens, it would no longer represent a rightful condition. Although there would still be the appearance of punishment, it is possible that such executive actions on the tyrant's part would not be true punishment, but rather merely the exercise of certain pragmatic rights of war against the citizens.

In addition to the need for a distinct executive, Kant also holds that the executive must be appropriately authoritative. Executive authority has two primary components.

---

[11] In making this claim, Kant largely adheres to the traditional position in favor of mixed governments. He does not fully utilize the familiar arguments of this view, however; instead, suggests that the sovereign and the executive cannot be the same person, as the executive is "put under obligation through the law of *another*, namely the sovereign" (6:317). This claim only explains why the two must be separate, though, if Kant is granted the additional, unstated premise that the executive must be under legal obligation. The legislator is already either a) under no obligation, or b) under its own obligation. It is not clear, just from Kant's statement, why the executive could not also be in either of these two conditions. While there are good reasons Kant could give for the necessity of keeping the legislative and executive separate, this one appears circular.

First, executive power must be wielded by an officially designated person or office. This might seem obvious, but it rules out the possibility of vigilante enforcement of the law as punishment. Only the legislative has the capacity to designate the individual or institution that will be responsible for imposing sanctions for the violation of law. Second, executive authority must be unified and strictly hierarchical. Kant thinks that there must be a single executive power, from which all other executive officials derive their authority. This is both practically necessary (to prevent conflict and civil strife) and conceptually necessary. Only a single executive figure will prevent an infinite regress of appeals from occurring (6:319).

As we have seen, Kant defines punishment as a coercive action, undertaken against a citizen of a state by the legitimate executive, as a sanction for the violation of public law. While a perfectly just state might not be required in order for these conditions to be met, they are only possible within a state that can, at the least, establish a rightful condition. Without such a rightful condition, there is no possibility for duties of right. In the next section, I fully explore the concept of these duties of right and their connection to punishment. As we will see, the very concepts of punishment and duties of right are inextricably intertwined.

3.2 Coercion and the Division of Duties

The *Metaphysics of Morals* is famously divided into two major sections: the *Rechtslehre*, or Doctrine of the Right, and the *Tugendlehre*, or Doctrine of Virtue. On one level, the distinction between the two is easy to understand: the Doctrine of Right

corresponds roughly with what we now call political or legal philosophy, and the Doctrine of Virtue corresponds to ethics. This simple description overlooks a number of complicating factors, including what is—for the purposes of this chapter—a key issue: duty. Both the Doctrine of Right and Doctrine of Virtue include references to duty, and each describes duty in a different way. Duties of right, or juridical duties, require us to perform—or prohibit us from performing—a specific action, and we are liable to be punished if we fail to satisfy these duties. On the other hand, ethical duties[12] are often – but not always – less specific about the particular actions that are required of us,[13] and we are not liable for punishment from the state if we fail to satisfy these duties.

There are different ways of understanding how Kant draws the division between these two kinds of duties, each based in particular passages from the *Metaphysics of Morals*. I will be arguing in favor of dividing juridical duties from ethical duties on the grounds of a duty's incentive: juridical duties allow for the possibility of external incentive, while ethical duties do not. Along with the other supporting reasons I will detail below, this reading of Kant is bolstered by the fact that this strategy for dividing the duties is the first one that he offers when introducing the concept of Right and Virtue. This strategy is in contrast with other traditional views that hold the division of duties to ultimately turn on the content of the duty. There are two possible kinds of

[12] The term 'ethical duties' is frequently used synonymously with 'duties of virtue.' I am avoiding using these two terms interchangeably. Following Paul Guyer, I will be using 'duties of virtue' to refer to a specific subset of ethical duties. See Guyer, *Kant*. 2nd ed. London: Routledge, 2014, p. 279.
[13] As we will see, it is a mistake to think that ethical duties are always more open-ended (i.e., that they are imperfect duties). Some ethical duties – specifically, perfect duties to oneself – require specific actions. More on this below.

content that might be relevant to this distinction, and I will address them each in turn. Finally, there is a fourth approach to the question of the division of duties, represented by Leslie Mulholland's work. According to this view, the appropriate way to separate juridical and ethical duties is to distinguish those duties that are associated with a correlative claim right from those duties that are associated with no such right.

The four possible positions on how Kant divides the duties—the one I will be defending as well as the three alternatives—all have a basis in Kant's own words. As such, rejecting any position will require me to either explain why the view misconstrues what Kant says or dismiss some of Kant's own claims as unsubstantiated. While the first strategy is preferable, it will be necessary at times for me to conclude that Kant has made a claim that he cannot support and that should be discarded. In these cases, I will be making such assessments in light of other passages in Kant's work that either directly contradict the problematic claims or are, at least, strongly in tension with these problematic claims. While we might be tempted to find a way to render all four views consistent and coextensive, this simply is not possible given the material in the *Metaphysics of Morals*.

Kant first addresses the division of the doctrine of right and doctrine of virtue in the general introduction of the *Metaphysics of Morals*. He writes in Section IV ("On the Division of a Metaphysics of Morals"),

> All lawgiving can…be distinguished with respect to the incentive (even if it agrees with another kind with respect to the action that it makes a duty, e.g., these actions might in all cases be external). That lawgiving which makes an action a duty and also makes this duty the incentive is *ethical*. But that lawgiving which does not include the incentive of duty in the law and so admits an incentive other than the idea of duty itself is *juridical*. It is clear that

> in the latter case this incentive which is something other than the idea of duty must be drawn from *pathological* determining grounds of choice, inclinations and aversions, and among these, from aversions; for it is a lawgiving which constrains, not an allurement, which invites.[14] (6:218-219)

Subsequently, he refines and clarifies this picture:

> The doctrine of right and the doctrine of virtue are therefore distinguished not so much by their different duties as by the difference in their lawgiving, which connects one incentive or the other with the law.
>
> Ethical lawgiving (even if the duties might be external) is that which *cannot* be external; juridical lawgiving is that which can also be external. (6:220)

In the two passages above, Kant specifies a decisive difference between juridical duties and ethical duties. Here, we see Kant suggesting that the distinction is based on the possibility of external motivation to satisfy the duty in the case of juridical duties and its impossibility in the case of ethical duties. The reason for a division of this sort comes from the nature of what the duties require of us. Juridical duties require persons to use—or refrain from using—their external freedom in a particular way, namely in accord with the universal principle of right. Ethical duties, on the other hand, require persons to use their internal freedom in particular ways; in other words, to will in accordance with maxims that satisfy the conditions necessitated by the categorical imperative. Failing to will in the way that duty requires does not infringe upon the freedom of others, and as such a failure to will the right maxim is not a violation of the universal principle of right. Only those actions whose omission would violate the universal principle of right can be coercively enforced.

---

[14] Kant's rejection of 'allurement' as a potential motivation for compliance with juridical obligations is not particularly well defended here. I will return to this passage in the third section of the chapter, wherein I will argue that Kant gives us no reason why allurement and a comprehensive system of rewards could not be instituted along with a system of punishment.

There is also a secondary, practical reason for limiting the use of coercion. The use of external force or coercion can be effective at bringing about compliance with juridical duties, but the same cannot be said with respect to ethical duties. Since ethical duties call for willing to be motivated by a particular maxim, no amount of external coercion can ensure that one's will is appropriately motivated. As we will see shortly, prioritizing this reason for the possibility of coercion, rather than considerations of external freedom, results in a different background justification for the division of duties.

According to the view that I am defending, then, the duties are distinguished by their possible incentives. While ethical duties can only be satisfied by one incentivized by duty itself, juridical duties can be satisfied simply by the performance of a specific action; the motive of the actor is irrelevant. This is not to suggest, however, that the agent who satisfies her juridical duties simply to avoid punishment acts with moral worth. Although this class of duties are 'satisfied' merely if we perform the requisite action, they are still Kantian duties. As such, the agent who satisfies her juridical duties out of respect for the moral law still acts in the only way that is characterized by moral worth. This position is supported by a number of Kant interpreters, including Mary Gregor,[15] Paul Guyer,[16] and Sharon Byrd and Joachim Hrushka.[17] Kant offers additional

[15] Gregor, Mary J. *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik der Sitten*. Oxford: Blackwell Publishing, 1963.
[16] Guyer, Paul. *Kant*. New York: Routledge, 2006.
[17] Byrd, B. Sharon and Hruschka, Joachim. *Kant's Doctrine of Right: A Commentary*. Cambridge: Cambridge University Press, 2010.

evidence for prioritizing the role that incentive plays in the introduction to the Doctrine of Right:

> Right need not be conceived as made up of two elements, namely an obligation in accordance with a law and an authorization of him who by his choice puts another under obligation to coerce him to fulfill it. Instead, one can locate the concept of right directly in the possibility of connecting universal reciprocal coercion with the freedom of everyone. That is to say, just as right generally has as its object only what is external in actions, so strict right, namely that which is not mingled with anything ethical, requires only external grounds for determining choice; for only then is it pure and not mixed with any precepts of virtue….Right and authorization to use coercion therefore mean one and the same thing. (6:232)

So far, I have presented quotations that clearly highlight the role that the possibility of external incentive plays in distinguishing juridical from ethical duties. These three passages are not the only ones that make this point, and I will be presenting more as I refute some of the alternative views of the division of duties. While I think that the incentive-focused view takes the best account of the whole of what Kant says, it is important to note that the above passages from the general introduction to the *Metaphysics of Morals* and introduction to the Doctrine of Right are not Kant's last word on the division of duties. In the introduction to the Doctrine of Virtue, Kant offers two related, yet distinct, explanations of how juridical and ethical duties differ. These two explanations correspond to the two content-focused alternative interpretations of how Kant divides the duties. I will present each of the content-focused views in turn, discussing how they relate to each other and why the incentive-focused view I have discussed is a better fit with the text than either of these alternatives.

The first passage from the introduction to the Doctrine of Virtue that grounds an alternative, content-focused view first appears in section II:

> [Ethics] cannot begin with the ends that a human being may set for himself and in accordance with them prescribe the maxims he is to adopt, that is, his duty; for that would be to adopt maxims on empirical grounds, and such grounds yield no concept of duty, since this concept (that categorical ought) has its root in pure reason alone. Consequently, if maxims were to be adopted on the basis of those ends (all of which are self-seeking), one could not really speak of the concept of duty. – Hence in ethics the *concept of duty* will lead to ends and will have to establish *maxims* with respect to ends we *ought* to set ourselves, grounding them in accordance with moral principles. (6:382)

He goes on to elaborate further in section VI:

> Only the concept of an *end* that is also a duty, a concept that belongs to exclusively to ethics, establishes a law for maxims of actions by subordinating the subjective end (that everyone has) to the objective end (that everyone ought to make his end). The imperative "You ought to make this or that (e.g., the happiness of others) your end" has to do with the matter of choice (an object). Now, no free action is possible unless the agent also intends an end (which is a matter of choice). Hence, if there is an end that is also a duty, the only condition that maxims of actions, as means to ends, must contain is that of qualifying for a possible giving of universal law. On the other hand, the end that is also a duty can make it a law to have such a maxim, although for the maxim itself there mere possibility of agreeing with a giving of universal law is already sufficient. (6:389)

These quotations, taken together, form the basis for what I will call the 'ends vs. actions' variety of the content-focused interpretation. According to this view, Kant distinguishes between juridical and ethical duties on the basis of what the duty requires of us. Juridical duties require us to perform or refrain from a certain action, while ethical duties require us to will a specific maxim or hold a specific end. According to this view, our juridical duties are determined by the universal principle of right, and the ends that are required by ethics are determined by the categorical imperative; we can only satisfy these obligations by being motivated to do our duty for the sake of the moral law.

There is a fair amount of overlap between the ends/actions view and the incentive-focused interpretation I am defending. For instance, it is clearly the case that one cannot be coerced into holding a specific end, and therefore any such duty to adopt

a specific end would belong to the Doctrine of Virtue. We can, however, be coerced into performing an action. On these grounds alone, then, it might appear that the 'ethical duties as ends' view is coextensive with – because it explains – the incentive-focused view. As we will see, however, there are noteworthy background differences between these two ways of dividing the duties, rendering it necessary for us to choose one or the other as the fundamental criterion for selection of duties into one group or the other. The ultimate criterion of selection will be the view that best fits with the body of Kant's work and best provides the theoretical tools to make sense of the system of punishment that he erects.

This view – that ethical duties are duties to hold a certain end – is defended by, among others, Allen Wood.[18] Wood begins his discussion of the division of duties by arguing directly against the position that I am defending. He offers what he sees as a compelling reason for thinking that the possibility of external incentive is not the correct site at which to draw the distinction between kinds of duties: not all duties of right are, in fact, coercible. He cites the so-called duties of equity and duties that a ruler owes to her people as examples of juridical duties that are nevertheless unenforceable and, therefore, incoercible.[19] He writes,

> [Kant] thinks the relevant duties of right [duties of equity and duties a ruler owes to his people] are valid even when there are no enforcement mechanisms. We therefore misunderstand the Kantian conception of "right" if we think of it as merely a philosophy of law and the state. Instead, right is a system of rational moral (*sittliche*) norms whose function is to guarantee the

[18] See Wood, Allen. *Kant's Ethical Thought*. Cambridge: Cambridge University Press, 1999, and Wood, Allen. *Kantian Ethics*. Cambridge: Cambridge University Press, 2008.
[19] Wood, 2008: pp. 161.

> treatment of humanity as an end in itself by protecting the external freedom of persons according to universal laws.[20]

Since there are duties of right that do not allow for the possibility of an external incentive as the motive for satisfying the duty, Wood argues, this cannot be the appropriate understanding of the grounds for the division of duties.

After making these negative arguments, Wood is left with the task of providing some kind of positive account of where we ought to draw the line between juridical and ethical duties. Part of the support for his positive account arises directly from his negative critique; if he has successfully demonstrated that the incentive-focused view is incorrect, then this gives more weight to the above passage from the introduction to the Doctrine of Virtue. In addition to these arguments, he offers a basis for his view by positing that ethical duties arise exclusively from the second formulation of the categorical imperative, the formula of humanity as an end in itself (FHE). The formula of universal law, he maintains, can only provide a formal "CI-procedure." In order to find any positive ethical duties, we must look to the FHE, which commands us to hold certain ends (i.e., our fellow human beings). He writes,

> The law that goes beyond the merely formal principle of duty has to do with the 'matter of choice,' namely with its *ends*. In other words, the foundations of a Kantian theory of ethical duties are *teleological*. The theory is based not on the inherent 'rightness' or 'wrongness' of actions but on which actions promote certain obligatory *ends* (our own perfection and the happiness of others).[21]

The ends that we are obligated to have are collected under the general heading of ethical duties, and they are ultimately reducible to obligations to treat the rational

[20] Ibid, 162.
[21] Ibid, 166.

personhood in ourselves or in others with love and/or respect. Juridical duties, on the other hand, are not obligations to hold an end of respecting humanity, but rather an obligation to act in a certain manner.

I will address Wood's negative arguments before moving on to his positive case. In the case of duties of equity, Wood seems to have misconstrued Kant's position. It is true that Kant addresses the subject of equity in the appendix to the introduction to the Doctrine of Right. It is even true that he says of equity that it "admits a right without coercion" (6:234). We should not, however, conclude anything from this quotation alone. To do so would be to miss the fact that Kant's aim in this appendix is to sort out the ambiguity that arises from the term 'right.' The idea of right to equity—or the right of necessity, which is also included in the same appendix—arises out of a 'wide' understanding of right, he says (6:234). This stands in contrast to the more narrow understanding that corresponds with law and the possibility of enforcement. Kant describes the ambiguity surrounding the two possible understandings of the term right in the following way:

> One sees that in both appraisals of what is right (in terms of a right of equity and a right of necessity) the *ambiguity* (*aequivocatio*) arises from confusing the objective with the subjective basis of exercising the right (before reason and before a court). What someone by himself recognizes on good grounds as right will not be confirmed by a court…for the concept of right, in these two cases, is not take in the same sense. (6:236)

It is not difficult to see why Wood reads this section as supporting certain juridical duties as incoercible and therefore unenforceable. If one takes Kant to be saying that the wider conception of right is correct, then Wood would be right. It is not clear from this passage or even the appendix as a whole, though, whether Kant is endorsing

the wider or narrower understanding of right. Given this ambiguity, I think there is good reason to take Kant as saying that the wider conception of right captures something, but not a juridical duty; only the narrower conception actually corresponds to duties of right. After all, Kant repeatedly emphasizes the connection between right and coercion throughout the *Metaphysics of Morals*. For example, he states, "What essentially distinguishes a duty of virtue from a duty of right is that external constraint to the latter kind of duty is morally possible, whereas the former is based only on self-restraint" (6:383). In this brief quotation, Kant describes duties of right as involving right in the narrow sense. In accepting Wood's view that Kant is supporting the wider understanding of right, we would be forced to reject all of the passages in which Kant explicitly endorses a more narrow interpretation of right. If we followed this route, we would be left without a clear definition of what Kant takes right to be or the nature of its connection to coercion. This seems like a far more foundational aspect of Kant's position than is its ability to fully address the nature of the 'rights' or equity and necessity. The preponderance of evidence, then, suggests that the appropriate way to understand the ambiguity surrounding the so-called duties of equity is as something other than duties of right. If they are not duties of right, then their incoercible and unenforceable nature poses no problems for the incentive-focused view I am defending.

If we accept that Wood is mistaken that duties of equity represent incoercible juridical duties, then he must rely on the ruler's obligation to her people to make his negative case. As we will see, though, this example is also not as straightforward as Wood presents it. Based on his citation, we can see that Wood reads Kant's *Theory and*

Practice as grounding the claim that a ruler has duties to his subjects that cannot be coercively enforced. First, it is important to note that there is a certain degree of ambiguity surrounding the term 'ruler.' Typically, when Kant speaks of a ruler, he has in mind the executive head of the state, rather than the legislative body.[22] Whether Wood means the executive or the legislative, however, it does not seem that he can point to any clear instance in which Kant describes authority as having duties directly to the citizens of the state. In *Theory and Practice*, Kant seems more concerned with working out the limitations of the state's legislative power. Indeed, he holds that any legislature that enacts a law that the citizens could rationally object to has overstepped its authority (8:297). This limitation, though, arises from what kind of power the state has been granted by the citizens, rather than specific duties owed to them. The state does not have the power because it cannot be given, not because the citizens exercise rights to constrain legislators.

Turning to the *Metaphysics of Morals*, however, we see that the idea of incoercible duties owed by an executive ruler to her people is even harder to defend. Given that Kant's discussion of the division of duties occurs within the *Metaphysics of Morals*, it makes better exegetical sense to look here for Kant's position on duties of a ruler to his subjects. Rather than taking no clear stand on the issue, as he does in *Theory/Practice*, here Kant clearly rejects the idea of executive obligations. "Now, from this principle follows the proposition: the sovereign has only rights against his subjects and no duties"

[22] For examples, see 6:316-317 and 6:320-321.

(6:319). The very reason why there are no duties owed by a sovereign to his subjects is that no one has the authority to coerce the sovereign into the performance of such duties.[23]

In the case of equity, then, we saw the Wood adopts a particular reading of an ambiguous section. There are no other passages that support his reading, but a significant number that support the reading that there are no unenforceable duties, such as equity. While it is not possible to say that his reading is obviously mistaken, the preponderance of evidence appears to tell against him. On the subject of a ruler's duty to his citizens, the outlooks are even worse. The closest Kant comes to endorsing anything like such duties is positing the view that the legislative body is limited in its authority and is only justified in passing laws of a certain form (i.e., those to which all citizens could hypothetically consent). It appears, then, that Kant is not committed to incoercible or unenforceable duties of the kind that Wood discusses. Without these unenforceable duties, Wood's negative case against the incentive-focused view collapses.

With Wood's negative case against the incentive-focused view undermined, we return to his positive arguments in favor of the view that duties are divided according to their content: specifically, ethical duties are duties to adopt a certain end. As we saw in the above passages, cited from the introduction to the Doctrine of Virtue, there is good reason to see Kant as making some version of this claim himself. Furthermore, without

[23] This is not to suggest that the executive authority cannot behave impermissibly. If she refuses to carry out the laws passed by the legislative body, then the executive has so behaved and can be replaced. This is not a wrong in the technical sense, though, as it is not one that can be enforced or punished, and as such Kant rejects the idea that she has violated a duty.

the negative cases as contrast between the incentive-view and the ends/actions view, the two positions begin to appear perfectly coextensive. Instead of interpreting the principle of division in a way that leads to significant differences in classification, these two positions end up appearing to do little more than focus on different aspects of the same division of duties.

Nevertheless, there are several reasons for thinking that the possibility of external incentive is still playing a more fundamental role, even if both methods of division produce seemingly identical results. The first reason for rejecting Wood's content focused view is that he construes all ethical duties as duties of virtue proper. While all duties of virtue are ethical, Kant is clear that "not all ethical duties are thereby duties of virtue" (6:383). Duties of virtue represent a specific subset of ethical duties in which the content of the duty is an end in itself. As Guyer writes, "The term 'duties of virtue' should be reserved for those duties that involve the promotion of the two necessary ends, but the term 'ethical duties' should be used for the broader class of all duties that may not be coercively enforced."[24] Wood, in other words, succeeds in dividing duties of virtue from all juridical duties and the broader class of ethical duties. If our aim, though, is to divide juridical duties from all ethical duties, then his account does not succeed.

Drawing this distinction more carefully allows us to better account for the otherwise perplexing presence of some of the duties that Kant offers in the Doctrine of

[24] Guyer, p. 279.

Virtue. Specifically, Kant includes in the Doctrine of Virtue certain duties that have the appearance of prescribing actions. The duty not to commit suicide and the duty to refrain from ridiculing others are two good examples that the ends/actions view cannot adequately handle.[25] These are still ethical duties, in that adherence to them is not coercively enforceable. They are not, however, duties of virtue.

The second reason for prioritizing the possibility of external incentive over the content of the duties in dividing juridical from ethical duties is purely textual. Wood has the passage from the introduction to the Doctrine of Virtue to support his view, but nowhere else does Kant explicitly endorse this position. Indeed, there is even a little evidence to suggest that he does not conceive of ethical duties first and foremost as duties to hold a specific end. He writes, "Ethical lawgiving (**even if the duties might be external**) is that which *cannot* be external; juridical lawgiving is that which can also be external" (6:220, emphasis mine). By raising the possibility of ethical duties whose content is external, Kant seems to be suggesting that there are, or can be, ethical duties that are not merely internal duties to hold a specific end.

In contrast, the incentive-focused view is supported in the general introduction, the Doctrine of Right, and even in the Doctrine of Virtue. Again and again, we see Kant making statements such as "The Doctrine of Right and the Doctrine of Virtue are therefore distinguished not so much by their different duties as by the difference in their lawgiving, which connects one incentive or the other with law" (6:220). He uses

[25] Guyer, pp. 277-286

descriptions of this kind when first introducing the concept of juridical and ethical duties, giving us additional reason to think that this means of dividing the duties is truly the most important and fundamental. Given the weight of the textual support, it seems that the incentive-focused position is simply the stronger of the two. Coupled with the fact that such a position distinguishes between duties of virtue and ethical duties generally and can therefore better account for some of the duties included in the Doctrine of Virtue, this preponderance of textual evidence carries the day for the incentive-focused view.

***

The next major, alternative reading of Kant maintains that the division between juridical and ethical duties lines up perfectly with another important division of duties: that of perfect and imperfect duties. This view, which I will call the perfect/imperfect view, holds that all duties of right are perfect[26] and all duties of virtue are imperfect. The categories of perfect and imperfect duties appear first in the *Groundwork of the Metaphysics of Morals* (henceforth, the *Groundwork*).[27] Perfect duties are those that we must satisfy, because their violation cannot be willed without contradiction; specifically, attempting to will the violation of a perfect duty results in a contradiction in conception, meaning that they are conceptually self-contradictory. Willing the violation of an imperfect duty is not self-contradictory in this way, but attempting to will in this way

[26] Specifically, this view characterizes duties of right as perfect duties to others.
[27] Kant, Immanuel. *Groundwork of the Metaphysics of Morals. Practical Philosophy*. Mary J. Gregor, Ed. Cambridge: Cambridge University Press, 1996.

still leads to a contradiction. In these cases, the contradiction is known as a contradiction in willing, and it arises out of a conflict between willing the immoral maxim and other ends that we all have as a result of the principle of rational willing.

Perfect and imperfect duties take on a slightly different character in the *Metaphysics of Morals*. Here, Kant seems less concerned with the question of the kind of contradiction that results from failing to act in accordance with the duty. Instead, he focuses more on the specificity with which the duty prescribes the obligation we are under. Perfect duties, in this case, take on the character of obligating us to perform a specific action; there is no room for interpretation, and he claims there will never be a conflict between duties of this sort. Imperfect duties, on the other hand, prescribe a more general obligation that could potentially be satisfied in a number of ways. There is potentially some tension between various imperfect duties, and as such it is up to us to determine in what way we will satisfy all of our different imperfect duties.

Proponents of the perfect/imperfect view maintain that these categories are coextensive with juridical and ethical duties. Once again, the basis for this reading is clearly articulated in the introduction to the Doctrine of Virtue. Kant writes,

> If the law can prescribe only the maxims of actions, not actions themselves, this is a sign that it leaves a playroom (*latitude*) the free choice in following (complying with) the law, that is, that the law cannot specify precisely in what way one is to act and how much one is to do by the action for an end that is also a duty….The wider the duty, therefore, the more imperfect is a man's obligation to action; as he, nevertheless, brings closer to *narrow* duty (duties of right) the maxim of complying with wide duty (in his disposition), so much the more perfect is his virtuous action.
>
> Imperfect duties alone are, accordingly, *duties of virtue*. (6:390)

Here, we see Kant rather clearly drawing the parallels between juridical duties and perfect duties on the one hand, and ethical duties and imperfect duties on the other. There is no doubt that, in this passage at least, Kant intends to demonstrate that these two different ways of dividing duties are equivalent or coextensive; there is, however, reason to doubt that this is what he should have said. As we will see, by making this parallel, he commits himself to the claim that all perfect duties are duties of right and all imperfect duties are duties of virtue. This claim is one that is contradicted by what he himself says elsewhere in the *Metaphysics of Morals* and in other published works. I will argue that there is good reason to believe his other statements, rather than to accept the proposed equivalence between the two kinds of division of duties.

Before examining this issue closely, we should pause to consider whether this reading is, in fact, a different view than the content-focused position that ethical duties are duties to hold ends. As I articulated in the introduction above, both of these two views take the content of a duty to be what determines whether it is juridical or ethical. Also, given that Kant lays the groundwork for the perfect vs. imperfect view immediately after spelling out the details of the ethical duties as ends view, it is reasonable to conclude that he meant these two criteria to be connected. Indeed, it is clear that in most cases, the two are directly related. The fact that juridical duties require or prohibit a specific action does not leave much room for interpretation or play; likewise, the fact that ethical duties require only that we adopt a specific maxim or end

tends to leave room for us to act in a number of possible ways while still fulfilling our duty.

To suggest that the perfect/imperfect view and the ethical duties as ends view are actually one position, however, would be to ignore important differences. For instance, one can imagine a wedge case that demonstrates the difference between the ends/actions view and the perfect/imperfect view. Consider perfect duties to self. As we will see, perfect duties to self –such as the duty not to commit suicide – are listed by Kant as ethical duties. The perfect/imperfect view cannot allow that there are these sorts of duties, whereas the ends/actions view conceivably can. In order to do so, Wood would merely need to argue that I do not truly satisfy my duty not to commit suicide if I desperately wish to do so but refrain merely from fear of the pain I will experience; satisfaction of such a duty requires me to hold myself as an appropriate kind of end. Given that these two views come apart in this case, it is reasonable to conclude that they are distinct and should be considered individually.

Having established the distinctiveness of the two content-focus views, I now turn to demonstrating the problems with the view that juridical duties are synonymous with perfect duties and ethical duties with imperfect duties. While my arguments against Wood and the ethical duties as ends view ultimately depends, at least in part, on a disagreement over how to interpret the textual evidence, the case against aligning the division of duties with the perfect/imperfect split is more straightforward. Specifically, a

few examples clearly demonstrate that this approach simply carves up the conceptual space incorrectly.

The first wedge case that demonstrates the problems with this view is that of perfect duties to self. The example of perfect duties to self that Kant offers in the *Groundwork* is the duty to refrain from committing suicide. Under this view, that would mean that the prohibition against suicide (and any other perfect duties to self) would be a duty of right. Such a classification, however, is textually contradicted by the catalogue of duties of virtue that Kant offers only pages later in the *Metaphysics of Morals*. The perfect duty owed to oneself not to commit suicide is literally the first kind of duty of virtue that he discusses. Given that this and other perfect duties to self are clearly included in the Doctrine of Virtue, it is fair to say that any strict equivalence between perfect/imperfect duties and juridical/ethical duties would require us to reject broad and explicit classifications on the grounds of a single passage. This kind of move would be interpretively irresponsible and thus cannot be maintained.

The second kind of case that indicates there are problems with the perfect/imperfect view is less clear cut, but still worth noting. Recall that the *Groundwork* introduces the distinction between perfect and imperfect duties as a matter of what kind of contradiction arises out of willing the violation of a given duty; willing the violation of a perfect duty results in a contradiction in conception, and willing the violation of an imperfect duty results in a contradiction in willing. Most cases of violence – those that are not in self-defense – are prohibited by juridical duties, not ethical ones. It does not

appear, however, that willing violence against another person truly results in a contradiction in conception. It might clearly lead to a contradiction in willing, but this would only make it an imperfect duty. As such, the perfect/imperfect view would classify this kind of duty as a duty of virtue, despite the fact that it is clearly a duty of right.[28] Focusing on the possibility of external incentive in determining how to classify a particular duty does not result in this problem.

Based on these objections, we can conclude that the perfect/imperfect view overlooks and is contradicted by important textual evidence. Although such a view might be grounded in the passage from the introduction to the Doctrine of Virtue, to take it seriously would require us to then dismiss the actual taxonomy that Kant offers only a few pages later. Even if we were tempted to do so, though, the issue of duties prohibiting violence would still spell trouble that the perfect/imperfect view cannot resolve. While it is seemingly true that all duties of right are perfect duties, it is not the case that all perfect duties are juridical or that all imperfect duties are ethical.

***

The third major view on what separates different kinds of duties from one another that I will be addressing focuses on a new feature of duties: claim rights. Instead of focusing on the incentive of the duty, as I do, or on the content of the duty, as Wood does and as a supporter of the perfect/imperfect position might, this third view is

[28] This argument is similar in nature, yet different in conclusion, than one made by Allen Wood. See Wood, 1999: pp. 98-100.

essentially relational and holds that duties of right are those that involve a correlative claim right, whereas ethical duties do not involve any such correlative right. This is the position developed and championed by Leslie Mulholland. As we will see, this is a robust view that nevertheless ultimately reduces to one of the preceding three positions. While I think that his position most plausibly reduces to a variation of the incentive-focused view, both of the content-focused views remain distinct possibilities for other versions of the claim rights view.

In describing his own position, Mulholland explains the nature of a claim:

> Claims about rights are made to insist that one person may not interfere with the actions or conditions of another, without the other having performed a deed that, because of a law, allows interference. This kind of demand cannot be made in a context governed by social interest or the common good….A system of rights requires that *rules* govern the interrelations of persons as equals, rather than promote the common good.[29]

A juridical duty, then, is one in which a citizen may be compelled to perform the duty by another citizen making a claim of the sort that is described above. An ethical duty, on the other hand, carries with it no correlative right; no one can insist that I carry out my ethical duties.

This manner of dividing duties appears at first glance to be quite distinct from the other three we have examined. When considering the question of why specific duties have claim rights associated with them, however, we begin to see that perhaps this approach is not as distinct as it first appears. After all, there are multiple ways in which

[29] Ibid, 141.

we might conceive of why some duties are associated with claim rights and why others are not.

In the first instance, we might posit that juridical duties are associated with claim rights in virtue of the fact that it is possible to insist of others that they behave in a specific way. There are no claim rights associated with ethical duties, in this case, because it is not possible to insist that others have a certain end. As we see, then, this is an approach to the claim rights view that ultimately grounds rights in something akin to Wood's ends/actions position. While others' actions are the kinds of things about which we can have claim rights, others' ends simply are not.

In the second instance, a variation of Mulholland's claim rights view might achieve its support by relying on the distinction between perfect and imperfect duties. In this case, the reason for the presence or absence of claim rights would be linked to the kind of obligation that the duty imposes on us. Given that a perfect duty specifies a particular action that ought to be performed, it is possible to make a claim in these cases; because imperfect duties prescribe only a more general disposition or group of possible actions, we cannot make a claim on another. If there is no clear act that an individual is obligated to perform, we cannot clearly claim a right that impels him to perform his duty.

Although the above reductions are possible for a claim rights view, there is reason to think that Mulholland's own view collapses into an incentive-focused view of

the kind that I am defending. In two passages that seem to support such a reduction, he writes,

> 1) To determine what rights persons have, it is necessary and sufficient to determine which wrong actions it would be right for others to coerce anyone to omit and which wrong omissions it would be right to coerce others to avoid.[30]

> 2) To mark off the class of duties corresponding to rights, Kant provides a test through which we can determine whether a duty is of this kind. Such a test will also be a test for rights. I put the principle which enables us to determine whether a person has a right as follows: A person has a right to something if, and only if, his having it or doing it is a condition under which the will of one person can be united together with the will of another in accordance with a universal law of freedom.[31]

In the first passage, Mulholland seems to clearly align his interpretation of Kant with the incentive-focused view. In the second passage, he goes further to connect his position with the interest in freedom that underlies and supports the incentive-focused view.

It might appear to be a strength that some version of the claim rights view is more or less consistent with the three different readings of Kant's division of duties that I previously discussed. However, we should avoid this conclusion. The ambiguity that enables the claim rights view to appear consistent with the other three alternatives also makes it difficult to determine what kind of fundamental commitments the position has. Although it ultimately appears that Mulholland is relying on the incentive-focused approach that I am defending, the fact that his position needs the support of another

[30] Ibid, 142
[31] Ibid, 143

indicates that dividing the duties on the grounds of their association with claim rights does not capture the fundamental difference between the kinds of duties.

All three alternative views, then, have significant textual or conceptual difficulties. Although they are largely grounded in passages from the *Metaphysics of Morals*, we should consider the text as a whole when deciding how to weigh any individual passage. When we attend to the full body of the work, it is clear that the preponderance of evidence supports the view that duties of right and duties of virtue are distinguished by whether or not they allow for a duty to be satisfied by an agent with a motive other than respect for the moral law.

3.3 Allurement and Reward

In the last section of this paper, I argued that the incentive of a duty is the appropriate grounds for deciding whether the duty belongs to right or to virtue. Ethical duties allow for no incentive other than the concept of moral law and duty itself. If we are motivated by anything else, we have failed to actually behave morally, thus failing to satisfy the obligation. Duties of right, on the other hand, can be satisfied when motivated by a wider range of possible incentives. Specifically, external incentives of aversion and allurement are admitted as permissible motivation.

Kant, however, is quick to argue that the only appropriate kind of external incentive is coercion. He writes,

> It is clear that in the latter case [duties of right] this incentive which is something other than the idea of duty must be drawn from *pathological* determining grounds of choice, inclinations and aversions, and among these, from aversions; for it is lawgiving, which constrains, not an allurement, which invites. (6:219)

His claim, then, is that aversive incentive is the only external form of incentive that can permissibly motivate us to the performance of duties of right. The reason for this claim, though, is not as clear. The only justification he offers in this brief passage is that the nature of lawgiving is inherently constraining, and as such it should not motivate us through allurement. It is my aim in this section to examine this claim. Ultimately, I show that there is no good basis for positing a strict difference between incentivizing by aversive coercion and by allurement. Instead, we should recognize that the state could justifiably offer rewards for the performance of duties of right, so long as there were still punishments in place should one fail to obey the law.

Kant's only argument on this point in his published works is the above quotation from the *Metaphysics of Morals*. Lawgiving, he says, is the kind of thing that constrains. We might reject his whole 'argument' by questioning this claim on its face; after all, the moral law is seen as no constraint on our will, as it flows from our own rational personhood. In the same way, the law is meant to flow from the rational personhood of the citizens of a state; no law is just if any citizen could object to it, qua rational citizen. As such, it is possible to envision a Kantian claim that the law of a state in the rightful condition is also not, by its nature, constraining. Rather, we are at our most free when we act in accordance with a law we give ourselves.

For the moment, however, let us accept Kant's view: there is an analytic connection between lawgiving and constraint. The particular way in which the law constrains is by creating new duties where none previously existed. In other words, it constrains by imposing new limitations on the ways in which we may use our external freedom. Even given that the law is inherently constraining in this fashion, though, it is not clear that this tells in favor of the conclusion Kant draws – that legal allurement is forbidden. In what way does the threat of punishment constrain? While a sanction might force compliance with the law, the means by which it constrains is by modifying the expected outcome of one's actions. In other words, the threat of punishment appeals to one's inclinations in order to constrain. This understanding of 'constraint' differs from the way in which law constrains. While the latter constrains by creating new duties and obligations that are consistent with reason – just as the moral law constrains – the former constrains by altering the projected costs and benefits of a given course of action. In this respect, reward and punishment are identical. Just like punishment, reward aims to modify the expected gains and losses of a given course of action so as to bring about a specific result. While we typically think of it as being easier to forgo a reward than to accept a punishment, this is merely a difference in degree, not a difference in kind.

We might reject these parallels, however, and maintain that law is analytically connected to punishment in an inseparable way. Juridical law, unlike the moral law, might be presumed to necessarily have an associated punishment, as a matter of definition. Even still, it is still not clear why the use of legal reward for compliance with the law should be prohibited. The necessity of a sanction for every law does not entail

that reward for compliance is impossible; all it entails is that reward *alone* cannot meet the criterion of lawgiving. Provided that there is still a system of punishment in place that provides the necessary constraint as a last resort, his argument gives us no reason to think that there could not be an additional scheme of rewards for successfully obeying the law. Allurement, in other words, could be an acceptable incentive for the performance of our duties of right, but it could not fully replace punishment as the sole means of enforcing the law.

Turning from Kant's published works, there are two passages from Vigilantius's notes from Kant's lectures on moral and political philosophy that indicate Kant was not always committed to the view that reward was an inadmissible motivation in the performance of duties of right. First, when discussing the various species of pathological motivation, he discusses two main categories:

> a. *per placentia, sive per illecebras*, though compulsion by something that pleases Is not in fact called compulsion; e.g., because it tastes so good.
> b. *per minas*, in regard to all disagreeable consequences. (27:522)

Nowhere in or around this taxonomy does Kant suggest that only the second kind of pathological motivation is appropriate for motivating us to perform our juridical duties. Rather, both are presented as essentially interchangeable with respect to their role in incentivization. At §38, however, he offers a similar taxonomy of the forms of pathological motivation, and this time he remarks on the different roles that these two forms play in motivation. In this second passage of note, he writes,

> The nature of duty does not allow of being coupled with the idea of reward. So reward can never be the motive of a moral act of duty, since the latter must be presented through the law itself. On ethical principles, therefore, an action

> undertaken in the hope of reward could never have morality, though it might well have legality. (27:548)

If we are not careful, we might read the first sentence of this passage as supporting the claim that reward is incompatible with the concept of law. This reading, however, is misconstruing the terms being used. In this passage, Kant is speaking of *ethical* duty; acting out of an interest in reward is incompatible with morality. He clearly states, however, that such a motivation is not at odds with legality. As duties of right are concerned with legality and not morality, then this passage is further evidence that at one time Kant conceived of reward as a possible motivation to perform duties of right. Given that he wrote the *Metaphysics of Morals* after the delivery of these lectures, it is possible that he changed his mind on this issue. If his only reason for changing his position was the above argument about lawgiving constraining, however, then my analysis of that argument shows we ought to give preference to his earlier views. He simply offers no substantial basis for ruling out allurement as an acceptable form of external incentive.

Even if we accept that rewards of this kind would be permissible, however, it is important to note that the state cannot rely solely on them, to the exclusion of punishment. Kant famously describes his goals for conceiving of an ideal political system as constructing a state that even a race of (rational) devils could inhabit peacefully and rightfully (8:366).[32] Even if devils might be inclined to satisfy duties of

[32] Kant, Immanuel. *Toward Perpetual Peace. Practical Philosophy*. Mary J. Gregor, Ed. Cambridge: Cambridge University Press, 1996.

right in the interest of receiving rewards, it is certainly conceivable that in some instances, one could benefit more from crime than from the reward for obeying the law. In such instances, any purely self-interested individual would have no reason not to violate the law. If our goal is to ensure that even these immoral but rational beings could exist in a rightful condition, then, we must include a system of punishment in order to prevent them from violating the rights of others.

We might ask, then, what the use of reward is. If we still need to have a comprehensive system of threats and punishments conjoined with all of our laws, what reason is there for rewarding those who follow the law? The answer to this question is two-fold. The first answer is the more minimal of the two: my aim in this section has been to show that reward is a possible external incentive for the performance of duties of right, not that we must or even should have such a system of rewards. In the passage I quoted at the beginning of the section, Kant states that allurement is contrary to the nature of law; it has been my aim to show that, at the very least, allurement is consistent with the law in the presence of a system of punishment.

I think, however, that we can show more than the mere permissibility of allurement. Indeed, there is good reason to think that Kant ought to find that the addition of some system of reward for obeying the law would be morally and politically desirable. To make this case, however, I will need to gesture toward material that will not be fully unpacked until next chapter.

According to Kant, the state has an interest in promoting freedom. While he does not consider the use of coercion against the freedom of wrong-doers problematic, Kant

would have to agree that the commission of crime is undesirable in light of the hindrance of the victim's freedom that it entails. It would be better from the perspective of protecting the freedom of citizens, all things considered, if fewer crimes were committed. A system of rewards for performance of one's juridical duties would help to accomplish exactly such an end. While some would undoubtedly still resort to crime and still receive punishment from the state as a result, the amount of coercion the state would need to impose would decrease dramatically. The threat of punishment is meant to reduce the commission of criminal actions, but the addition of a second source of deterrence would only improve legal compliance. In addition, punishment represents two impositions on freedom: the crime violates the victim's freedom and is wrong according to the universal principle of right, and the punishment of the criminal, although consistent with the universal principle of right and justified by its role in deterring future hindrances of freedom, is nevertheless still a hindrance of freedom itself. With a comprehensive system of rewards in place, both of these hindrances of freedom could be avoided. The would-be criminal would refrain from hindering the victim's freedom out of an interest in being rewarded, and therefore the state would not be required to hinder the criminal's freedom through the imposition of coercive punishment.

As a final point in favor of the use of a system of reward to incentivize the performance of one's duties of right, Kant makes some noteworthy psychological claims in Collins's lecture notes. He writes that, compared with the threat of punishment,

> Rewards are in better accordance with morality, since I do the action because its consequences are agreeable, and will be able to cherish the law which promises me reward for my good deed; but I cannot so love the law which threatens punishment. Love, however, is a stronger motivating ground for doing the action. (27:288)

Although following the law out of hope for reward is no more a moral motivation than doing so out of fear of punishment, Kant takes it to be a psychological truth about human beings that constraining our actions in order to avoid punishment will eventually give rise to feelings of resentment toward the law itself (ibid). This phenomenon will not occur, however, when we constrain our actions to secure a reward; instead, such a case will cause in us feelings of gratitude toward the law. Resentment will make it harder for us to obey the law in the future, whereas gratitude will make it easier. Although this psychological claim might be open to empirical investigation and rejection, it at least indicates that Kant saw legal reward as not only permissible, but even desirable.

Conclusion

Kant's definition of punishment is well incorporated into his fundamental positions in practical philosophy. It reflects both his deep division of the spheres of moral and political philosophy, as well as the ways in which the latter is founded upon the former. By defining punishment as only possible as the product of an institutional arrangement within a rightful condition, Kant builds in strict limits to what counts as an instance of punishing. This enables him to identify a very specific phenomenon as punishment, which in turn makes it much easier to construct a coherent, unified theory

of punishment. As we will see in the next chapter, this specificity will be both a cause of innumerable difficulties and the source of the solution.

While we have reached an understanding of what Kant takes punishment to be, it still remains to be seen why he thinks that the government has the power to engage in the use of coercion of this kind. This question – the justification of punishment – will be the focus of the next chapter.

4 Justifying Punishment: Deterring Threats to Freedom

Why does a state have the authority to punish its citizens? Immanuel Kant famously answers this question in a starkly retributive manner: the only permissible grounds for punishing is the immediate response to a prior act of wrongdoing, and justice is best served when the penalty resembles the crime as closely as possible (6:331).[1] Indeed, his strong support for a retributive justification for punishment has had such a sizeable influence on the history of punitive theory that Kant's name has been, for much of the past two centuries, nearly synonymous with backward-looking, retributive punishment. The textual evidence that Kant held this view seems overwhelming; the available support for it within Kant's practical philosophy, on the other hand, is deeply lacking. The few arguments Kant does provide demonstrate neither that retributivism is necessarily the state's justification for punishment, nor that this justification would even be consistent and compatible with the fundamental, distinctive elements of his practical philosophy. In short, Kant cannot coherently defend a retributive theory without abandoning one or more significant tenets of his practical philosophy.

For Kant, the state is neither simply a means of securing individual rights set forth by transcendent laws of nature, nor merely formed according to the conditional, pragmatic choices of the self-interested contracting agents. The state is necessitated by

[1] Kant, Immanuel. *Metaphysics of Morals*. Trans. by Mary J. Gregor. Cambridge: Cambridge University Press, 1996. All internal citations refer to the standard Prussian Academy edition, volume and pages.

contingent facts about the world we occupy,[2] but given this, its nature is partially fixed by the moral personality of its citizens. The state can create new rights and obligations, but it must do so in accordance with the basic moral obligations that individuals owe to one another, *qua* rational agents. Whatever justification Kant can offer for punishment, it needs to ground a theory that can reasonably subsist within his overarching conception of the purpose and role of the state. As I will argue, retributivism does not and cannot fit within this conception.

Rather than abandon the possibility of a truly Kantian theory of punishment, [3] I contend that Kant has the necessary resources to construct a theory that relies on deterrence for its justification. Specifically, the state's function of determining and preserving the external right of its citizens requires the active prevention of future violations of such freedoms. At the same time, certain retributive constraints – such as a prohibition against punishing the innocent – can be incorporated as practical policy guidelines for the implementation and application of this deterrent institution. This approach, which closely resembles a Hart-style 'mixed theory,'[4] has both the ability to fully cohere with Kant's practical philosophy and the virtue of preserving many of Kant's explicitly endorsed claims regarding the nature and requirements of punishment.

---

[2] For instance, we occupy a planet with finite space, limited resources, etc. Had we occupied a different kind of world or possessed different physical forms, the need for a civil state might be diminished or eliminated.

[3] This is the conclusion supported by Murphy, Jeffrie G. "Does Kant Have a Theory of Punishment?" *Columbia Law Review*, Vol. 87, No. 3 (Apr., 1987), pp. 509-553.

[4] Hart, H.L.A. *Punishment and Responsibility*: *Essays in the Philosophy of Law.* 2nd ed. Oxford: Oxford University Press, 1970.

I am not alone in arguing that a mixed theory that ultimately relies on deterrence for the justification of the institution of legal punishment is the most feasible and internally consistent option available to Kant. This school of thought has gained increased prominence among Kant's interpreters in recent years, owing to dissatisfaction with the arguments he gives and the tension they sense between retributivism and some of the most distinctive elements of Kant's moral theory – in particular, the Formula of Humanity.[5] The contours of Kantian deterrence were shaped largely by Sharon Byrd's influential paper on the role of deterrence and retribution in Kant's theory of punishment.[6] Her argument – that the state's interest in preserving itself provides a deterrent basis for the threat of punishment, which is subsequently carried out in a retributive manner – has inspired significant respect and a healthy following among those who study Kant's practical philosophy. Arthur Ripstein features prominently among these followers, albeit with several slight adjustments.[7] Although there are many sub-varieties of Kantian deterrence, those who justify punishment by reference to the state's right or duty to preserve itself occupy a prominent position.

---

[5] Examples include Hill, Thomas E., JR. "Kant on Punishment: A Coherent Mix of Deterrence and Retribution?" *Annual Review of Law and Ethics* 5, (1997), pp. 291-314 and "Treating Criminals as Ends in Themselves." *Jahrbuch fuer Recht und Ethik*, 11 (2003): 17-36; Holtman, Sarah. "Toward Social Reform: Kant's Penal Theory Reinterpreted," *Utilitas*, 9 (1997): pp. 3-21; and Merle, Jean-Christophe. "A Kantian Critique of Kant's Theory of Punihsment." *Law and Philosophy*, Vol. 19, No.3 (May, 2000), pp. 311-3338; Scheid, Don. E. "Kant's Retributivism." *Ethics*, 93 (1983), pp. 262-282; and Wood, Allen W. *Kantian Ethics*. Cambridge: Cambridge University Press, 2008.

[6] Byrd, B. Sharon. "Kant's Theory of Punishment: Deterrence in Its Threat, Retribution in Its Execution." *Law and Philosophy*, Vol. 8, No. 2 (Aug., 1989), pp. 151-200.

[7] Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge: Harvard University Press, 2009.

Despite broad areas of agreement, I find this tradition of Kantian deterrence to be unsatisfying. By framing punishment as necessary for the preservation of the state, Byrd, Ripstein, and others expose Kantian deterrence to a number of difficulties. As I will show, by routing the justification for punishment through the preservation of the state, this standard approach to Kantian deterrence views all crimes as equally significant, diminishes the value of individual citizens – victims and criminals alike – and loses sight of the original purpose of the state. My own approach – Kantian protective deterrence – avoids these problems by highlighting the value of individual autonomy and the state's role in protecting each citizen from any and all infringements of her or his right to external freedom.

4.1 Kant on Retributivism

Despite Kant's staunch commitment to retributivism, he spends surprisingly little time defending it. The standard accounts of Kant on punishment usually highlight either some Kantian conception of moral desert or Kant's belief that a deterrent system of punishment would necessarily use criminals unjustifiably as a means to an end.[8] As I read Kant, both moral desert and concerns about using persons as means play important roles in his justification, functioning together as interlocking arguments. His argument supporting moral desert is positive – meaning it is meant to demonstrate that retributivism is not only acceptable but morally required – while the second argument,

[8] See, for instance, Murphy, Jeffrie G. "Kant's Theory of Criminal Punishment." *Retribution, Justice, and Therapy: Essays in the Philosophy of Law*. (Dordrecht, Holland: D. Reidel, 1979), pp. 82-92.

concerning the use of criminals as a means, is purely negative – meaning it attempts to show that no other justification for punishment can be morally permissible. While Kant's positive argument fails on the grounds that it runs afoul of his foundational divide between the doctrine of right and the doctrine of virtue, his negative argument is more successful. On its own, however, it does little to show that retributivism is justified. Indeed, it is even possible that the constraint imposed by this argument could be accommodated by non-retributive theories of punishment.

In his first argument, Kant makes the positive claim that punishment must be retributively justified. He does this by connecting the authorization to punish and the determination of the appropriate amount of punishment with the inner moral worth of the character of the criminal agent. Thomas Hill calls this view the "intrinsic desert thesis," and I will follow suit.[9] According to the intrinsic desert thesis, "Acts of certain kinds have as an intrinsic property that it is *fit, appropriate, or 'called for' that the perpetrator suffer for it*…It takes no moral argument but merely conceptual analysis or moral intuition to 'see' that immoral…acts make the agent intrinsically deserving of painful sanctions."[10] Kant's first argument is intended to demonstrate the appropriateness of the intrinsic desert thesis. I call this argument the 'inner wickedness' argument. Kant writes,

> This fitting of punishment to the crime, which can only occur by a judge imposing the death sentence in accordance with the strict law of retribution, is shown by the fact that only by this is a sentence of death pronounced on every criminal in proportion to his *inner wickedness (*even when the crime is

[9] Hill, Thomas E., Jr. *Human Welfare and Moral Worth*: *Kantian Perspectives*. Oxford: Oxford University Press, 2002, pp. 310-362.
[10] Ibid., p. 325

> not murder but another crime against the state that can be paid for only by death). - Suppose that some...who took part in the recent Scottish rebellion believed that by their uprising they were only performing a duty they owed to the House of Stuart, while others on the contrary were out for their private interests; and suppose that the judgment pronounced by the highest court had been that each is free to make the choice between death and convict labor. I say that in this case the man of honor would choose death, and the scoundrel convict labor. This comes along with the nature of the human mind; for the man of honor is acquainted with something that he values even more highly than life, namely honor, while the scoundrel considers it better to live in shame than not to live at all.... Since the man of honor is undeniably less deserving of punishment than the other, both would be punished quite proportionately if all alike were sentenced to death; the man of honor would be punished mildly in terms of his sensibilities and the scoundrel severely in terms of his. On the other hand, if both were sentenced to convict labor the man of honor would be punished too severely and the other too mildly for his vile action. And so here, too, when sentence is pronounced on a number of criminals united in a plot, the best equalizer before public justice is death. (6:333-6:334)

Here we find Kant explicitly claiming that when considering the appropriateness, the proper amount, or the correct method of punishment, taking into account the inner character and motivation of the criminal is not only acceptable, but even necessary to ensure that justice is done. This view is consistent with statements he makes in his lectures on ethics, in which the inner wickedness argument also finds support. He is recorded by Vigilantius as saying, "In punishments, a physical evil is coupled to moral badness. … This link is a necessary one, and physical evil a direct consequence of moral badness" (27:552).

The idea that punishment is warranted by – and its proportion determined in accordance with – inner wickedness contradicts the nature of the relationship between the state and morality that Kant has previously established. The state exists in order to make possible the free, moral interactions of human beings (6:312), but it does not have a

direct role in shaping the character of its citizens (8:290-291).[11] There are several reasons for this: assessing another's moral character is virtually impossible, changing another so as to bring about a more moral character is almost as difficult, and even if these things could be accomplished, entrusting the state with the kind of power is dangerous.

Kant accounts for this by specifying the state's realm of concern as the actions of its citizens. He does this by dividing right and virtue. The basis for this division lies in the permissibility of enforcement of one set of duties (duties of right, or juridical duties), but not of the other (duties of virtue, or ethical duties). Failure to perform duties of right may be punishable, but failure to perform duties of virtue is not. Another central component of the distinction between these two classes of moral duties is that an agent's success or failure to satisfy duties of right is entirely independent of considerations about her or his motivations, maxims, or character. One with a wicked character could obey the law and be unpunishable, just as one with a good character could violate the

[11] There is a partial exception in the case of educating children. Kant does have an account of the proper approach to moral education in children (See Kant, Immanuel. *Education*. Ann Arbor: University of Michigan Press, 1960). For obvious reasons, this account involves more than mere rational discussion and appeals to the child's humanity. Various other factors, from the use of examples, the expression of disapproval, and the cultivation of obedience and a "longing to be honoured and loved" (ibid., p. 87). These methods of educating, however, are only permissible because Kant views children as significantly distinct from autonomous adults. Indeed, he expressed doubt that adults are capable of being trained by the same mechanisms as children (Buchner, Edward Franklin. *The Educational Theory of Immanuel Kant*. New York: AMS Press, 1971, p. 268). Further, Kant's writings all describe education as a process that occurs between parents, tutors, or governesses and children; it is likely that he would have resisted the idea of the state taking over as the primary educator. Nevertheless, the case of the moral education of children might still be instructive. Assuming that his concerns about state-guided education could be solved by mediating education through particular teachers, then it is clear the state can, and indeed should, aim to instill certain moral values. It is only permitted to mold the character of children because they are not fully rational, autonomous citizens, however; once a person reaches adulthood, such efforts would be both ineffective and an intrusive violation of freedom. I will return to these questions in chapter six, in which I discuss rehabilitation.

law and, in retributive language, deserve punishment. Arguably, one's character is determined more by her or his performance of duties of virtue than right, and failure to perform these duties is never sufficient to allow for the use of external coercion or force. Even a basic understanding of the concept of juridical duties shows that punishment cannot be based upon the inner wickedness of the criminal. The state's right to coercion does not extend to the character of its citizens. Punishment is not a response to the moral character or intrinsic desert of the criminal. Rather, punishment is a response to unacceptable and impermissible *behavior*.

One might try to find such a basis in Kant's writings on the subject of the highest good. Kant's view of the highest good is typically described as a condition in which people enjoy happiness in proportion to their virtue, and ideally they would experience a perfect degree of both (5:110–111). There is some debate, however, over the question of whether the converse of this principle is also true of his highest good; should people also experience suffering in proportion to their vice? If so, then bringing about the alignment between viciousness and the suffering of the wicked, as a component of the highest good, could justify the state's responding to character and moral desert.

This view, however, has two major problems. First, it is not clear that attaining the highest good requires the suffering of the wicked; the preponderance of scholarship on Kant's conception of the highest good suggests there is no such linkage.[12] Even if we were to grant the interpretation of the highest good that requires the suffering of the

[12] For example, see Silber, John "The Importance of the Highest Good in Kant's Ethics." *Ethics*. Vol. 73 (1962), pp. 179-197; and Wood, Allen W. *Kant's Moral Religion*. Ithaca: Cornell University Press, 1970.

vicious, however, this would still not serve as a satisfactory justification for punishment. The reason for this is simple: even if attaining the highest good does require that the vicious experience suffering for their wicked characters and deeds, there is no indication that it is the job of the state to bring about this congruence. This is what I refer to as the 'hard problem' of retributivism. In order to account for the state's authorization to respond to the moral deservingness of some to suffer, a theory must explain how such a permission fits within the overall purpose of the state. According to Kant, the state's role is limited to the determination and preservation of each citizen's innate right to their exercise of her or his external freedom, as well as whatever rights they can legitimately acquire through interactions with other citizens. It exists in order to establish and preserve a set of conditions in which citizens can experience a maximally reciprocal amount of personal freedom. It does this through the enforcement of duties of right. Although this arrangement – known as a rightful condition – is a necessary precondition for human beings to achieve the highest good, it is not itself the highest good. In other words, the state does not exist to address defects in the character of its citizens. Kant is quite clear about the fact that any government behaves illegitimately if it attempts to legislate based upon what will produce the greatest happiness or good for citizens (8:290-291). Likewise, the state has no authorization to attempt to bring about the highest good by attempting to arrange it such that every citizen enjoys happiness and suffering in proportion to her or his virtue and vice.

In light of the failure of the inner wickedness argument, Kant's support for the intrinsic desert thesis is left without a sound basis. Consequently, without the intrinsic

desert thesis, Kant does not offer a justification for retributive punishment; he never explains why the state is authorized to retroactively respond to moral desert with coercive force. Even if we allow for the existence of moral desert as analytically contained in the concept of immoral activity, and thus that wrongdoers *deserve* to be punished, this does nothing to demonstrate that the state can or should respond to such desert. Kantian retributive punishment, then, remains wholly unjustified.[13]

Kant's second, negative argument, on the other hand, is much more consistent with both his more foundational practical philosophy and his other statements on punishment. In this case, however, he intends merely to show that any other justification would be morally unacceptable. This argument, which I will be calling the 'second formulation' argument, makes its central claim on the strength of Kant's second formulation of the categorical imperative, the so-called Formula of Humanity. This argument appears in several different places in Kant's writings, but perhaps its clearest statement reads,

> Punishment by a court…can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society…. For a human being can never be treated merely as a means to the purposes of another or be put among the objects of rights to things: his innate personality protects

[13] Samuel Fleischacker tries to develop an account of Kantian retributivism that relies not on the analyticity of wrongdoing and desert for suffering, but rather on the Formula of Universal Law (FUL). (Fleischacker, Samuel. "Kant's Theory of Punishment." *Essays on Kant's Political Philosophy*. Chicago: University of Chicago Press, 1992, pp. 191-212). According to his approach, criminals should be made to suffer the results of the universalization of their maxims. Thus, the criminal who steals should be made to live in a world without the possibility of property, meaning that he or she will be reduced to surviving as a slave of the state. It does not seem clear, though, that all crimes would produce a reasonable punishment when the result of the universalization of its maxim is applied back to the criminal. Furthermore, this account seems like a very different version of proportionality than Kant describes elsewhere. The most significant problem for Fleishacker's theory, though, is that it still does not answer the question of why the state is authorized to make criminals experience the universalization of their maxims. As long as this 'hard problem' of retributivism remains unanswered, there remains a gap between desert (or impermissible maxims) and punishment.

> him from this…. He must previously have been found *punishable* before any thought can be given to drawing from his punishment something of use for himself or his fellow citizens. (6:331)

Although the last line indicates that some non-retributive side-effects of punishment are acceptable, punishing with only a deterrent or rehabilitative justification in mind is not, as doing so would be using the criminal as a means to achieve some other end. Our moral duty to respect the rational personhood of the criminal requires us to avoid using him or her as a mere instrument to accomplish some other end. In deterrent cases, criminals are used as a means to prevent future crime; in rehabilitative cases, the freedom of criminals as rational persons and ends in themselves is violated in the interest of improving the character or lives of the criminals themselves. Both of these might be admirable goals, in a sense, but to act on them is to fail to take seriously the value of the individuals being punished. In using criminals to accomplish social goals, the state reduces the value of free, rational beings to the instrumental value of a thing. The only way we can treat criminals as an end in themselves, Kant says, is to punish them only on the basis of their prior act of wrongdoing—in other words, retributively.[14]

It is worth noting, however, that the fact that a criminal's punishment generates some deterrent force – that is to say, causes the criminal herself or others to be less likely to commit a similar offense in the future – is not a problem for Kant. He is fully prepared to recognize that such an effect is likely to follow from acts of punishment (6:331). As long as generating such effects is not the aim or purpose of the punishment, however,

[14] In the third section of this chapter, I will show why deterrent justifications for punishment need not necessarily run afoul of the FHE the way Kant anticipates.

then the criminal is not treated as a means. Deterrence, then, is an acceptable by-product of punishment, but cannot be a justification.

Unlike the inner wickedness argument, the second formulation argument is internally coherent and well-supported by more foundational elements of Kant's moral philosophy. It cannot be ignored or rejected; to do so would be to suggest that the state need not follow the most basic of Kant's moral laws. Any successful interpretation of Kant's theory of punishment needs to explain how the institution of punishment can avoid treating criminals merely as a means to an end. Despite its strength, however, the second formulation argument is still limited in what it can show. It works well as a negative argument – it imposes a constraint that any interpretation of Kant's theory of punishment must accommodate – but it cannot actually *justify* the existence of punishment as an institution on its own. Kant still needs some positive account of why punishment is acceptable.

The positive and negative arguments are meant to work in tandem. The inner wickedness argument is intended to show why retribution is justified, and the second formulation argument shows why nothing else can take its place. Removing the positive argument, however, leaves the theory inert, unable to actually provide a reason why the state is justified in using coercive force against its citizens. In some abstract, moral sense, those who behave wrongly might deserve unhappiness, in the same sense as those who are virtuous deserve to enjoy happiness equal to their virtue. But this desert means little if Kant cannot provide any argument for why the state is entitled to answer these deserts.

This is the motivation for exploring the possibility of a deterrence justification. The very idea of Kantian deterrence might sound necessarily contradictory, and most of these efforts are pursued explicitly as constructive, rather than strict interpretation. There is, however, good textual reason to think that, at one time, Kant saw state punishment as justified by deterrent interests. In the lecture notes recorded by Collins, Kant writes,

> All punishments belong either to the justice or the prudence of the lawgiver. The first are moral, the second pragmatic punishments. Moral punishments are imposed because a sin has been committed…. Pragmatic punishments are imposed so that a sin shall not be committed; they are a means of preventing crime…. **All the punishments of princes and governments are pragmatic, the purpose being either to correct or to present an example to others. Authority punishes, not because a crime has been committed, but so that it shall not be committed**. But every crime, in addition to this punishment, has a property of deserving to be punished, because it has taken place. Such punishments, which must therefore necessarily follow upon the actions, are moral in character. (27:286, emphasis added)

As a rule, we should be careful about allowing material from the lectures to trump Kant's positions in published works. Kant takes a clearly retributivist position by 1788 with the publication of the *Critique of Practical Reason*,[15] and no later writings ever support the view from the lectures. Further, even though the Collins quotation appears to rule out the possibility of retributive punishment by states (the fact that wrongdoing generates moral desert is a precondition for the permissibility of state punishment, but it cannot justify state punishment), it is merely posited, with no arguments given to indicate how Kant thinks such a view should be supported. This passage, however, can

[15] Although Kant does not discuss punishment or even political or legal philosophy at length in the *Critique of Practical Reason*, he does allude to moral desert and retributivism. See 5:37.

still serve as a useful as a window into what Kant takes to be necessary for a retributive justification to succeed. Namely, retributive punishment must be executed by a moral being, whose interest in punishing is explicitly a response to moral desert. Given that Kant never argues the state is a moral person in the traditional, natural law manner, and given that he has contrasted this with pragmatic, deterrent punishment, it would appear that the only option left to him is to justify punishment by reference to its deterrent value.

4.2 Standard Kantian Deterrence

While it might be possible to construct a more stable foundation for Kant's retributivism,[16] the majority of the constructive efforts in Kant scholarship have centered on the possibility of a Kantian deterrent theory of punishment. This is due, in significant part, to Sharon Byrd, whose paper on Kant's theory of punishment[17] is referenced by most subsequent supporters of some version of Kantian deterrence. It seems appropriate, then, to address the 'standard' deterrence readings of Kant by focusing, at least in part, on her position. The hallmark of her approach to interpreting Kant is her division of punishment into a legal threat of sanction and the subsequent execution of

[16] Arguably, the most promising way of defending Kant's retributivism would be to construct a more contractarian account of his political philosophy. If the state could be shaped purely by what the contractors agree to, and one could successfully make the argument that agents would only agree to a state that punished for retributive reasons, then the result would be a deeply retributive theory of punishment. Both the first and second premises, however, would be very difficult to impossible to reconcile with Kant's other foundational political commitments.

[17] Byrd, 1989.

the sentence.[18] According to Byrd, the punishment for legal wrongdoing is threatened on the grounds that to do so deters citizens from committing crimes; once a crime has been committed, however, the punishment is carried out for retributive reasons.[19] This way of dividing up Kant's theory of punishment is not altogether dissimilar from my own approach, Kantian protective deterrence. As I will argue in the next section, the deterrent value of the threat of legal sanction is the best justification for the application of punitive coercive force that Kant has available.

The difference between Byrd's interpretation and Kantian protective deterrence is located in how she justifies the state's authorization to deter crime. She begins her account by focusing on the special status of the state: namely, that its existence is necessitated as a precondition for other moral ends. This grounds both the state's right to force individuals to enter into and remain in civil society, as well as the individuals' obligation to do so.[20] This duty to enter civil society is grounded on the necessity of property ownership. Briefly stated, she skillfully interprets Kant in the following way: while it is possible to have real possession in the state of nature, ideal possession (i.e., ownership without physical detention) cannot be possible, as there is no means in the state of nature by which to obligate others to refrain from using the remotely-owned

---

[18] Ibid., pp. 152-153

[19] Ibid., p. 153

[20] The duty to enter into states is a strange sort of duty. On the one hand, it cannot be a duty of right, as such duties are only possible in civil society. On the other hand, if duties of virtue cannot be coercively enforced, then it does not seem as though the duty to enter a state can be ethical, as entrance into the state can be coerced. The solution to this apparent puzzle is that the duty to enter into states is indeed a duty of virtue, and specifically, a perfect one. The right the state has to force individuals to enter, however, is not correlative with this duty; rather, the state's right is entailed by the state's right to self-protection. In other words, when the state forces an individual to join, it is not enforcing the individual's obligation to enter civil society; rather, it is only defending itself from a potential threat.

object without violating their freedom in the process. Yet, Kant establishes the possibility of ideal possession as a synthetic a priori truth: although we do not know it is possible analytically, we can come to know its possibility by recognizing that ideal possession is necessary for the existence of human free choice (itself known to be possible and, indeed, existent). Given that ideal possession is both possible and necessary for human free choice, we are obligated to bring it about. In order to do so, we must first bring it about that a state exists and we live as members of it.[21]

Given that the state alone can perform this special function, it is necessary for states to preserve themselves by eliminating any and all threats to their existence. Crime, according to Byrd, is necessarily a threat to the continued existence of the state.[22] As such, the state has not only an interest in, but also an obligation to prevent crime from occurring. The deterrent force generated by the application of punishment to criminal wrongdoing is precisely the means of accomplishing such prevention.

She argues that crime is a threat to the continued existence in two different ways, each of which warrants taking steps to minimize criminal behavior. The first way in which crime threatens the state is very literal and tangible: when citizens commit crime, they are lost as members of the state. She lays out her argument in the following passage:

> Society's right to punish criminal violations lies ... in its duty, as an expression of the common will, to maintain itself. Since one can force every other with whom one comes in contact to leave the state of nature and move to civil society, by the same reasoning one can also prevent anyone from leaving civil society and returning to the state of nature. Kant refers to

[21] Ibid., pp. 173-180
[22] Ibid., p. 181

> commutative justice in the state of nature as the 'condition of war.' In moving from the state of nature to civil society one subjects oneself to the common will which legislates and judges. Its judgments are enforce-able through legitimate force to secure the rights it decides upon. Only distributive justice in civil society can maintain peace and universal freedom.
>
> Crimes are violations of law that 'make the actor incapable of being a citizen.' Only public crimes are 'criminally punishable,' and public crimes are those that endanger the security of society.'[23]

Her argument is straightforward and clear. The state has a duty to maintain itself, which entails a related duty to prevent its citizens from leaving for civil society for the state of nature, as this would bring about the dissolution of the state. Further, the commission of a public crime results in the loss of the criminal's civil personality. If the state does not prevent crimes from occurring, then it will be allowing citizens to leave the state, by way of their criminally-induced loss of civic personality. As such, Byrd thinks the state is authorized to threaten the use of coercive force as a way of preventing such an eventuality from coming to pass.

Justifying deterrent punishment by reference to its ability to keep citizens from engaging in behavior that will strip them of their citizenship is a very indirect approach to the issue. Kant does describe criminal violations of law as those that make the perpetrator unfit to be a citizen (6:331), but to suggest that preventing this loss of civic personality is the justification for punishment is to lose sight of the other effects of an act of crime. Every crime involves the infringement on the rightful freedom of one or more citizens. The state has a duty to maintain itself because it is meant to prevent such infringements. Justifying punishment by reference to its duty to maintain itself, rather

[23] Ibid.

than to punishment's ability to accomplish the state's end of preserving the freedom of its citizens, introduces an unnecessary, extra step into the theory.

In addition, there might be deeper difficulties awaiting this strategy. For instance, it is not clear that preventing individual citizens from losing their citizenship is of great importance to Kant. After all, he describes a citizen as one who is fit to vote (6:314). In order to be fit to vote, one must be independent; Kant famously excludes 'dependents' such as servants from voting, as they lack the necessary independence to truly exercise their own will, free from economic compulsion. While there are compelling readings of Kant that suggest the state's economic policy ought to strive for granting independence and therefore citizenship to all subjects,[24] Kant does state that any degree of inequality in wealth is consistent with justice (8:291-292). Further, it would seem as though the loss of citizenship that attends criminal action would be identical in the relevant characteristics to a subject's loss of citizenship when he or she went from being a black smith to a woodcutter (6:314).[25] Presumably the latter does not justify the state in interfering with the citizen's freedom, so it is hard to see how the loss of citizenship alone would justify punishment.

[24] See Williams, Howard. "Toward a Kantian Theory of International Distributive Justice." *Kantian Review*. Vol 15, no. 2 (2011), pp. 43-77; and Wood, pp. 193-205.

[25] According to Kant, the black smith is independent because she owns the products of her labor. The woodcutter, on the other hand, does not; rather, he is hired to cut wood for a client, and thus sells his time. This indicates that he relies on others for his livelihood in a way that the blacksmith does not. There is plenty of room to question this distinction, but for our purposes it suffices to show that Kant conceives of many subjects of the state as lacking citizenship and that an individual could lose her or his citizenship by transitioning to one of these occupations.

Arthur Ripstein stakes out a position that is explicitly allied with Byrd's approach, and in particular, this second kind of argument she puts forward.[26] Like Byrd, Ripstein's aim is to defend a reading of Kant that holds punishment to be both retributive and deterrent. Also like Byrd, he posits that the way of defending such a view is by means of a justification for punishment that depends on the value and necessity of the state. He writes,

> Deterrence and retribution are united through Kant's view of punishment as something that can only be done by a superior…. I will argue that the criminal exempts him- or herself from public law, and is liable to punishment simply because public law cannot permit unilateral exemptions. Punishment is the guarantee that public law is effective in space and time…. The threat of public law is…the announcement that public law will remain supreme.[27]

Ripstein differs from Byrd, however, on the question of *why* the state is valuable. For Byrd, the state serves an instrumental role: namely, it makes possible ideal ownership, which is itself necessary for our exercise of free choice.[28] Outside of securing the prerequisites for right, the state has no intrinsic value. Ripstein disagrees with this view.[29] For Ripstein, the state is good in itself, as its own end:

> To characterize something as a means or instrument suggests that it serves to achieve something that might exist apart from it. Where Byrd writes of means or instruments, I will argue that Kant posits an identity: civil society *is* the systematic realization of freedom, required *a priori*, 'however well disposed and right-loving human beings might be.' In turn, the criminal law is an integral part of civil society, for it is nothing more than the supremacy of public law against opposing individual wills, should there turn out to be any. The enforcement of its prohibitions is itself equivalent to the prohibitions themselves.[30]

---

[26] Ripstein, pp. 302-303.
[27] Ibid., p. 302
[28] Byrd, pp. 153-154.
[29] Ripstein, p. 303
[30] Ibid.

This description of the value of the state is right on at least one point: the purpose of the state could not be achieved in any other way. I believe Byrd would readily accept this. Where Ripstein goes wrong, though, is thinking that the necessity of a specific means to an end somehow makes the means no longer instrumental. Although Ripstein might be right that the state is constitutive of a rightful condition, the rightful condition is still only a means to an end. In order to illustrate this, imagine a world in which rational beings are completely independent and self-sufficient; they do not need one another to achieve whatever ends they set, and they are incapable of harming one another. In such a world, the state would be neither necessary nor valuable; indeed, it would represent only an impermissible constraint on the wills of the rational beings. This demonstrates that the state is not a thing of unconditional value – freedom is. While the state might be the only moral way to secure freedom for all, this does not mean that the state is constitutive of freedom.

Setting this issue aside, however, we see that the core of Ripstein's interpretation of Kant's theory of punishment is very similar to Byrd's. According to Risptein, law allows for a community of separate persons to live together while retaining their independence; it does so by enabling the possibility of giving laws to ourselves.[31] This is only possible through the institutions of a state.[32] When an individual commits an act of crime, she wills in such a way so as to contradict both a particular law and the ominlateral, general will that created the law. Given that law is by definition

[31] Ibid., p. 231
[32] Ibid., p. 309

authoritative, and crime threatens this authority, the state is justified in attempting to prevent such violations from occurring.[33] The fact that crime infringes on another's right is of no immediate significance; only its contradiction of the authority of law is.

Ripstein articulates this view explicitly, citing the example of theft:

> The ground for *punishing* theft, however, is not the fact that the thief chooses to violate the basic norm of property. Instead, the grounds for punishment reflect the fact that his choosing to do so must be understood as choosing to exempt himself from the authority of the law.[34]

Although Byrd does not explicitly make the same claim, her arguments align with Ripstein's. By focusing on the way in which any and all crime threatens the nature of law and the authority of the state, Byrd and Ripstein both justify the deterrent application of punishment by highlighting the preservation of the state. Crime is understood only abstractly, as a threat to the state, stripped of all its particular details. As we will see, this account of crime cannot successfully justify punishment in any recognizably Kantian manner.

It might be tempting to respond to Byrd and Ripstein by claiming that the maxim of an individual criminal is no real threat to the state's authority. After all, Kant writes in the *Groundwork*,

> If we now attend to ourselves in any transgression of a duty, we find that we do not really will that our maxim should become a universal law, since that is impossible for us, but that the opposite of our maxim should instead remain a universal law, only we take the liberty of making an *exception* to it for ourselves (or just this once) to the advantage of our inclination. (4:424)

[33] Ibid., p. 313
[34] Ibid.

Impermissible maxims are not those that seek to destroy the power of the state; instead, they merely seek to exempt the criminal from its authority. This response cannot succeed, however, because it is precisely this exemption that endangers the state and its authority. As Byrd and Ripstein point out, the very concept of public law requires that it apply to all persons, without exception. By making exceptions of themselves, criminals pose as great a threat to the possibility of law as they would if they sought actively to undermine it.

This does not mean, however, that such a strategy can successfully serve as the justification for punishment. Holding that the state is justified in using punitive coercive force solely in order to maintain its own authority necessarily leads to serious obstacles for a Kantian theory of punishment. If crime is wrong because it threatens the state then all crime is essentially identical. Whether one vandalizes a building, steals from his neighbor, or murders a fellow citizen, the crime is punishable because it threatens the state's supremacy and authority. This is especially true in Ripstein's account, in light of the non-instrumental value he attributes to the state. Given that he views the state as good in itself, he cannot even reference the ways in which a particular crime wrongs individuals or undermines the aims or ends of the state. Rather, he can only explain why crime is impermissible by reference to it as a threat to the existence of the state. As we saw above, this is a matter that he recognizes and explicitly endorses.[35]

[35] Ibid.

The blanket uniformity of crime under the 'preservation of the state' style deterrence theories has several strange or undesirable outcomes. First, we must totally abandon Kant's insistence that punishment be fitting or proportional to the crime committed. Although there are problems with a strict application of *ius talionis*, the importance of some relatively fixed proportionality between crime and punishment is a central theme in Kant's theory of punishment. If our concern is truly to examine or develop a fully Kantian theory of punishment, we ought not reject proportionality unless absolutely necessary. Ripstein's arguments for deterrence that focus on the preservation of state authority necessitate the abandonment of proportionality between crime and punishment, but do not give any direct reasons for why this is an interpretively acceptable move.

Perhaps Byrd or Ripstein might try to get around this problem by suggesting that the state could still punish proportionally based on the degree of threat that an action posed to the authority of the state. This proposal, however, cannot get off the ground; both Byrd and Ripstein are clear that the reason a crime threatens the state's authority is that it involves the willing of a maxim by an individual that undermines the omnilateral nature of public law. In this way, a simple act of trespassing is as great a threat to the state's supremacy as is a killing spree. Both ignore the nature of the law in the same way, and as such, both pose equal threats to the concept of law.

A second peculiar outcome of the Byrd/Ripstein position is that it seems to lose sight of the relationship Kant envisions between the state and the individual citizens.

Kant describes the rightful condition as arising entirely out of the concept of the external freedom of the individuals who seek to bind themselves together in a civil society (8:289). The laws that they create impose certain limitations and constraints on their freedom, but as these constraints arise from the wills of the citizens, they each remain free and autonomous. Establishing and preserving these conditions is the purpose of the state; any time it deviates – either by allowing some too much freedom at the expense of others or creating laws that abridge citizens' freedom in ways they could not consent to (8:297) – it not only fails to achieve its purpose, but it actively works against it. The state, then, is not only an instrument to make possible the exercise of external freedom, it is one that is only morally justified when it functions properly.

It seems natural, then, to justify the prevention of crime in light of the state's role in establishing and preserving the conditions that allow for citizens to live together freely. According to the 'preservation of the state' style arguments for deterrence, however, preventing instances of crime is only of indirect interest. Instead, proponents of this approach seek to justify punishment primarily as a means of preserving the status quo. Although both Byrd and Ripstein would likely argue that the continued existence of the state is necessary for it to fulfill its role in preserving the external freedom of the citizens, their position involves an unnecessary, extra step. Rather than justifying punishment as a means of preserving the state, which in turn allows for the protection of individuals' external freedom, why not simply justify punishment as necessary for the protection of individuals' external freedom?

None of the outcomes of arguments based on preserving state authority is internally contradictory or untenable; a perfectly consistent theory of punishment could include such tenets. Attributing them to Kant, however, seems highly questionable. While Byrd and Ripstein are undoubtedly correct that every violation of law is a challenge to the authority of the state, to conclude that this is the sole justification for punishment is to radically alter Kant's conception of the limited state. Preserving the authority of the state is necessary, but it is necessary for the sake of the state's role in protecting the external freedom of the citizens. If punishment already accomplishes this – as I hope to show below – then it seems as though punishment's effects on reinforcing the supremacy of the state are more of a happy side effect than the justification. Rather than a necessary instrument for coexistence that determines obligations via law and enforces sanctions against those who violate such laws, the Byrd/Ripstein Kantian state becomes an apparently self-justifying moral entity, whose citizens seem to exist so as to perpetuate the state.

In summary, the traditional approaches to Kantian deterrence have relied on very abstract arguments to justify the institution of punishment. These arguments have focused on the state and its legal authority, claiming that preserving these institutions and their supremacy can serve as a basis for punishment. Although Kant clearly holds that the state must preserve itself, to ground punishment in this obligation is to ignore key facts about the institution of punishment within the broader context of the state. Taken to its conclusions, this approach robs our ability to distinguish between crimes or

punish them in a way that is proportionate and fitting. Additionally, the importance of individuals and their right to the free exercise of their external freedom play only an indirect role for either Byrd or Ripstein. As I will show below, it is both possible and preferable to construct a theory of Kantian deterrence in which protecting the external freedom of the citizens is the primary force driving the justification of punishment, while preserving the state and its authority plays only an incidental role.

4.3 Kantian Protective Deterrence

If Kant offers no substantial basis for retributivism, and the existing models of deterrence face their own obstacles, one might be tempted to conclude that Kant lacks the resources to explain the permissibility of the state's use of punishment.[36] According to this way of thinking, Kant has structured his basic political philosophy in such a way as to contradict the requirements of his moral philosophy; in turn, punishment (at least defined in the way that Kant does) is necessarily unjustifiable in a Kantian state. We ought to reject this conclusion, however, on the grounds that there is a viable, deterrent solution to this problem that remains firmly grounded in the fundamentals of Kant's practical philosophy. I call this solution Kantian protective deterrence. Despite similarities to Byrd and Ripstein's views, Kantian protective deterrence distinguishes itself by justifying the institution of punishment in an individual-focused, forward looking manner. In short, the state is justified in threatening punishment on the grounds

[36] This argument is made in full by Murphy (1987).

that each act of crime contravenes the purpose of the state by violating a particular citizen's right to the free exercise of her or his external freedom. Punishment is then carried out as a necessary means of preserving the efficacy of the threat of legal sanctions. Finally, Kantian protective deterrence avoids using criminals as a means by highlighting the role that rational consent plays in the Kantian legal framework.

In order to understand how this particular application of coercive force is grounded, it is first necessary to discuss how coercion of any sort is justified. As the section progresses, we will trace Kant's argument from the innate right to freedom to the reasons for the foundation of the state, and finally we will arrive at coercion's crucial role in grounding punishment as a necessary institution to preserve these reasons for the state's existence.

Conceptually, Kant's first discussion of coercion pertains to individuals in the state of nature. It is worth noting, that Kant did not take the state of nature to have been a literal, historical condition of humankind; rather, any discussion of people in the state of nature is a thought experiment designed to specify which kinds of rights arise purely from personhood and which arise only from the interactions of people as citizens of a state.[37] The right to act coercively in the name of self-defense is, according to Kant, a right of the former kind. Vigilantius records Kant as stating that "Everyone can resist the freedom of another, so soon as it infringes that freedom of his own, which is able to co-

[37] For a clear discussion of what Kant means by the "original contract," see Byrd and Hruschka, 2010: pp. 170-171.

exist with the freedom of everyone else" (27:524). He also describes the right to self-defense in the *Metaphysics of Morals*: "It is not necessary to wait for actual hostility; one is authorized to use coercion against someone who already, by his nature, threatens him with coercion" (6:307). This indicates that even in pre-civil interactions, individuals act permissibly when they use coercion in to defend themselves from acts of hostility.

Although the right to self-defense might seem basic, Kant offers an explanation for it: the innate right to freedom that is held by all rational beings. According to Kant, this is the only right that exists independently of states. He describes the innate right to freedom in *Theory and Practice* (8:290), but his most relevant description of the right comes in the Doctrine of Right:

> *Freedom* (independence from being constrained by another's choice), insofar as it can coexist with the freedom of every other in accordance with a universal law, is the only original right belonging to every man by virtue of his humanity. (6:237)

The connection between every individual's right to universalizable external freedom and coercive self-defense is also presented as analytic. Kant's definition of a right, essentially, is the title to use coercion to bring about whatever it is the right entails (6:230). When another acts to limit our freedom against our wishes, her actions fail to accord with universal law. As such, the innate right to freedom we all possess justifies us in the use of coercion in defending ourselves, even when such self-defense occurs at the expense of the aggressor's freedom. I would even be justified in using coercive force to protect another from aggression by a third party, should I see such an attack transpiring or be aware of one to come imminently.

There is a significant difference, however, between the right to personal or even collective self-defense against an aggressor and a state-enforced system of punishment, carried out against violators after the fact. Kant cannot claim that the right to punish is grounded solely in our use of self-defense; after all, the execution of punishment is carried out by other individuals, after a wrong has been committed. Unlike Locke, Kant does not claim that individuals have an unlimited executive right in the state of nature; the right to punish is one that can only exist within the state.[38] In order to get from coercion in self-defense to coercion in state instituted punishment, then, Kant will need additional arguments.

He supplies these missing steps by articulating the purpose of the state and its role in preserving and promoting freedom. Kant famously argues that all individuals who must interact with other human beings are under an obligation to enter into a formal state. He writes, "From private right in the state of nature there proceeds the postulate of public right: when you cannot avoid living side by side with all others, you ought to leave the state of nature and proceed with them into a rightful condition, that is, a condition of distributive justice" (6:307). As we saw in Byrd's account of state authority, this obligation is a result of the state's unique capacity for creating exclusive rights to objects in the world.

[38] Murphy, Jeffrie G. *Kant: The Philosophy of Right*. Macon: Mercer University Press, 1994. Pp 95-107.

Although the state is the only means of guaranteeing the possibility of freedom through enabling the ideal ownership of property, its purpose is not merely limited to establishing the conditions that allow for property rights. Rather, the state has a motivating interest in protecting and promoting the external freedom of its citizens by guaranteeing that everyone behaves in a manner consistent with universal right.[39] Each action that is not consistent with universal right represents the illegitimate abridgment of a citizen's external freedom and represents a failure on the state's part to achieve its purpose. All of its ends can ultimately be traced to establishing the conditions that will enable as perfect a protection of external right as possible. This protection of freedom takes two connected forms. First, the state must clearly articulate the kinds of behaviors that can coexist with one another under the universal principle of right; in other words, it makes determinate the obligations of citizens toward one another. It does this through enacting law, prescribing and prohibiting actions that will enable everyone to coexist in a rightful condition. The second aspect of the state's protection of freedom is the other purpose of law: specifying sanctions, or the negative consequences that will result from refusing to use one's freedom in a way that is consistent with the freedom of all other citizens.

It still remains for me to demonstrate how Kantian protective deterrence functions; in other words, how can a system of punishment preserve freedom by

[39] Guyer, 2006: pp. 279-281.

deterring individual acts of criminal wrongdoing? While discussing the state's authorization to punish, Kant analytically connects the state's role in preserving and promoting freedom with punishment:

> Resistance that counteracts the hindering of an effect promotes this effect and is consistent with it. Now whatever is wrong is a hindrance to freedom in accordance with universal laws. But coercion is a hindrance or resistance to freedom. Therefore, if a certain use of freedom is itself a hindrance to freedom in accordance with universal laws (i.e., wrong), coercion that is opposed to this (as a hindering of a hindrance to freedom) is right. Hence there is connected with right by the principle of contradiction an authorization to coerce someone who infringes upon it. (6:231)

It is clear from this passage that Kant takes coercion to be justified in certain cases simply by the law of double-negation. If hindering a person's freedom is bad, preventing someone from hindering another's freedom is good. Accepting this double-negation, however, does not get us all the way to a justification for punishment. The argument in this passage can go so far as authorizing direct intervention to prevent the hindering of a citizen's freedom; if the police were to witness a crime in progress, they would be justified in stepping in to prevent the hindrance of the victim's freedom. The argument, however, does not explain why the police would be justified in bringing a criminal to justice after the commission of a crime has been completed. After a crime has been fully committed, the act of punishing the criminal represents the use of a state's power to coercively limit the freedom of the criminal; in other words, the state actively promotes a hindrance to a citizen's freedom. If the state's fundamental purpose is to protect freedom, such a hindrance seems difficult to justify. It seems then, that any coercion after the fact (i.e., punishment) cannot be justified on the same grounds as the

interference in a present, on-going hindrance to freedom. It does not follow analytically from the state's role in protecting freedom – and therefore hindering hindrances to freedom – that the state is also permitted to issue punishment to the offender.[40] Law as a system of threats is exactly the missing piece of this puzzle.

It might seem psychologically obvious that being punished for violating the law is an effective way at decreasing the likelihood that individuals will repeat the same offense. It might even seem obvious that such punishing will discourage others from committing the same offense. Both of these assumptions, however, depend on the condition that both the criminal and other observers know the reason for the punishment. In other words, deterrent effect depends upon an open and publicly announced causal link between a specific act of wrong-doing and some form of coercion. From here, it is only a short step to the concept of law and sanction; instead of merely announcing the link between crime and punishment after one has done wrong, the state declares such a link ahead of any particular instance of wrong-doing and reliably enforces it following any and all criminal actions.

[40] Bernd Ludwig attempted to solve this problem simply by reference to the universal principle of right; see Ludwig, Bernd. *Kants Rechtshlehre*. *Kant Forschungen*, bd. 2. Hamburg: Meiner, 1988. Pp. 96-98. He argues that we can distinguish the freedom of the victim from the freedom of the criminal based on one's compatibility with the universal principle of right and the other's incompatibility. This idea is fairly straightforward; if the freedom of the criminal is being used in a way that is not rightful, then such freedom need not be preserved. We saw Kant making this very claim in his justification for self-defense. While this move is consistent with retributive readings of Kant, it does not actually solve the problem of why punishment is justified after a crime has been committed. It still gives us only a justification for intercession, rather than a fully-fledged institution of punishment.

This solution provides Kant with a method of claiming that punishment is justified as the hindering of a hindrance to freedom, albeit in something of an indirect manner. By declaring that any hindrance of freedom that does not accord with the universal principle of right will be met by a specific punishment, the state can effectively deter crime before it occurs and thereby promote freedom. The actual act of punishing itself would not, in most cases,[41] actually hinder any hindrances to freedom. Instead, punishment is necessary in order to support the system of threats that actually deters crime. If we did not punish, no one would believe the threats, and thus the laws would lose their deterrent efficacy.

This move is akin to Byrd's division between the threat of punishment and the execution of punishment. It is the threat that deters crime, and so the threat is what is directly justified by the state's forward-looking interest. But without actual sanctions, threatening legal sanctions would not give potential criminals any reason to refrain from illegal action. The sanctions, then, become instrumental, useful in order to insure the efficacy of the threat. One might raise concerns over this kind of account; it seems as though in punishing criminals, the state uses them as a means to achieve the deterrence force of the law. This concern, however, is defused by Kant's requirement that law be

[41] The only scenarios where this would obviously be the case would be situations in which the punishment itself necessarily prevents any future wrong-doing. Executing is an extreme example of this, but imprisonment (especially incarceration for the remaining duration of one's life) can also be thought of as actively hindering any future hindrances to freedom. In both cases, however, we would need to know with certainty that the individual in question was going to commit crime in the future. Given the impossibility of this kind of knowledge, execution or incarceration might, potentially, be an act of coercion enforced on an individual who would never again hinder another citizen's freedom.

willable by all citizens. If the state chose individuals at random in order to maximize the deterrent efficacy of the law, this would clearly be using citizens merely as a means. The law, however, specifies that only those who have engaged in crime may be used in this fashion.[42] This arrangement is something that any rational citizen could accept, and as such the state essentially has the permission of the citizens to use them in such a manner. Thus, although the state uses criminals as a means to achieve deterrence, it does not use them *merely* as a means, as it has their rational consent.

Unlike Byrd and Ripstein's positions, though, Kantian protective deterrence is squarely focused on deterring crime in order to prevent future violations of the citizens' right to freedom. While the more abstract interest in preservation of the state and its supreme authority are also accomplished by Kantian protective deterrence, they are secondary ends. This might seem like too narrow an account of crime; after all, there are many illegal actions that do not directly interfere with any other citizen's external freedom. Whether in cases of printing counterfeit money (6:331) or in an instance of trespassing in which no damage is caused, it is not clear that the perpetrator has actually violated another's freedom. If I cannot explain the illegality of cases such as these, this would be a serious problem.

I do not think that this objection poses a real threat to Kantian protective deterrence. According to Kant, public laws are specified in order to establish the

[42] This allows Kantian protective deterrence to capture much of the retributive character of Kant's theory, yet contextualize it within a theory justified by deterrence.

boundaries of each individual's external right. Even if I do not physically harm anyone or damage any property when I break a law, I still act in ways that infringe upon one or more persons' right. The state has an interest in preventing such behavior, as it must protect the system of right, even if no individual citizen lodges a complaint. This response is still distinct from Byrd and Ripstein's views, as they would highlight the incompatibility of the maxim with the state's authority, whereas Kantian protective deterrence focuses on the incompatibility of the maxim with others' rights to external freedom as specified by a system of public law.

In addition, Kantian protective deterrence has the advantage when it comes to questions of proportionality. Unlike Byrd and Ripstein's 'preservation of the state' approach, which cannot explain why we ought to punish different crimes with different sanctions, my position can easily account for proportionality. Given that Kantian protective deterrence focuses on the individual violation of one or more others' rights involved in instances of crime, it can easily justify the application of sanctions of varying severity. An act of trespassing is a much less serious infringement of another's right than is an act of murder, and the punishment should reflect this. The ability to capture this key element of Kant's theory of punishment is one of the major strength of Kantian protective deterrence.

Conclusion

Kant's reputation as a retributivist is grounded in the indisputable: despite minor deviations in his lectures, all of his published writings explicitly and consistently defend the view that punishment is justified solely as a response to a prior act of wrongdoing. The primary arguments he gives to support this position aim to show that – what is known as the intrinsic desert thesis. Additionally, he argues, any other justification for punishment would necessarily involve using the punished party as a means to some other end, whether it be deterrence or rehabilitation.

These arguments, however, are irreconcilably inconsistent with various elements of Kant's most basic positions in moral and political philosophy. Even if we accept that wrongdoing analytically entails a desert to suffer, Kant lacks the means to show that the state is justified in answering this desert. In light of these difficulties, some interpreters of Kant propose reading Kant as offering – or able to offer – a deterrent theory of punishment. The predominant manner of justifying such a deterrent theory has involved the state's duty to preserve itself and the omnilateral supremacy of law, both of which are threatened by the willing of any illegal action. This approach, however, cannot distinguish between different kinds of crime, and it runs the risk of making the effects of crime on individual citizens of only secondary or indirect significance.

Instead of founding punishment on the state's duty to preserve itself, I propose a deterrent justification that focuses primarily on preventing violations of the citizens' right to the use of their external freedom. This justification is supported by the state's

role in establishing the conditions that allow for individuals to live together, each enjoying her or his innate right to freedom. By threatening punishment, the state actively hinders unjust hindrances to freedom and thereby fulfills one of its primary purposes. As we will see in the next chapter, however, this picture of deterrent punishment at times appears at odds with other explicit statements Kant makes with respect to punishment. Although there are significant challenges, Kantian protective deterrence is capable of reconciling most of Kant's starkly retributivist statements with the deterrent justifications that underlie the most coherent interpretations of his theory of punishment.

5 The Liability of Punishment: All and Only the Guilty

In the last chapter, I argued that Kant's commitment to retributivism is not supported by sound arguments. Instead, the most stable basis he has for constructing a full theory of punishment rests upon a deterrent justification. Even though Kant's arguments in the Doctrine of Right fail to support a retributive justification for punishment, however, I do not contest that Kant himself was a staunch retributivist. Although there are a few telling passages in his lectures that indicate that, at least in his earlier years, he was possibly comfortable with a deterrent justification for punishment, there are numerous statements in his published works that are unambiguously retributive in character. We need only to look briefly at his most complete, published writings on the subject to see confirmation of this; Kant writes, "Punishment can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society. It must always be inflicted upon him only *because he has committed a crime*" (6:331).[1]

What ought my interpretation – Kantian protective deterrence – to do with statements of this kind? If these numerous claims are merely the products of an unworkable commitment to retributivism then the most obvious thing to do would be to reject everything Kant wrote on the subject of punishment. According to this approach, all of his statements about punishment are hopelessly corrupted by the fact that he relies

[1] Kant, Immanuel. *Metaphysics of Morals*. Trans. by Mary J. Gregor. Cambridge: Cambridge University Press, 1996. All internal citations refer to the standard Prussian Academy edition, volume and pages.

on an unstable and unsupported justification. To conclude, however, that everything Kant wrote on the subject of punishment is simply in error – conflicting irreconcilably with more fundamental aspects of his political philosophy – would place Kantian protective deterrence very much at odds with Kant himself. As I stated in the introduction to this dissertation, I am interpretively committed to conserving as much of Kant's published position as possible, and so this kind of wholesale dismissal would represent a serious failure.

While such outright rejection might be necessary in a few cases, in this chapter I will show that the majority of Kant's retributive statements can be retained and accommodated within Kantian protective deterrence. Rather than rejection of the problematic statements, our approach ought to be one of reconciliation and rehabilitation. Specifically, I will show how it is possible to preserve many of Kant's most retributive passages by understanding them as referring not to the justification of punishment, but rather to its liability. Recall that, according to my conception of what constitutes a theory of punishment, a theory can pick out various features of how the institution of punishment will operate that remain partially distinct from the justification. While they ultimately must accord with the justifying aim of punishment, they are not merely reducible to this aim. Among these elements is a specification of who should be punished.

According to the traditional, retributive interpretation of Kant's theory of punishment, he selects as liable all and only those who have done wrong. In other words, he establishes a fully retributive liability. This selection is apparently made on

the basis of moral desert (6:333-334). To punish one who has not done wrong would be to act impermissibly, just as the state would act impermissibly if it failed to punish one who has wronged. If our aim is to preserve as many of Kant's expressed views as possible, then these two conditions ought to be accommodated – Kantian protective deterrence ought to embrace a retributive liability, in which all and only those who have done wrong are punished. The basis for this liability, however, will have to differ from Kant's own; rather than relying on moral desert, I will need to supply other considerations, in keeping with the deterrent justification of punishment. I will accomplish this with a mixture of principled and empirical arguments.

This chapter contains three major sections. First, I will make the case that many of Kant's most retributive sounding passages can be reinterpreted as referring to liability, rather than justification. Further, I contend that adopting a retributive liability in this manner is perfectly consistent with a deterrent justification for punishment, provided that we use the correct understanding of a 'mixed' theory of punishment. Second, I will argue that Kantian protective deterrence can easily show that only those who have done wrong may be punished; the formula of humanity – coupled with the Kantian conception of legislation – prevents the punishing of innocents, regardless of any deterrent efficacy that such an act of punishment might generate. Third, I will address the more difficult subject of punishing all of the guilty. On the one hand, I will argue that Kantian protective deterrence has the means to support a policy of punishing all wrongdoers. On the other hand, I must concede that it cannot quite capture the full force of Kant's claim that all of those who have done wrong must be punished.

Nevertheless, Kant does not effectively make a case for this claim and that his most compelling example is beset with numerous problems.

5.1 Liability and Mixed Theories

Upon closer examination, many of Kant's most famously retributive passages do not directly refer to the justification of the state's use of punitive coercive force. Take, for instance, his claim that, "The *right to punish* is the right a ruler has against a subject to inflict pain upon him because of his having committed a crime" (6:331).

He expresses a similar view on the law of retribution, *ius talionis*:[2]

> Accordingly, every murderer – anyone who commits murder, orders it, or is an accomplice in it – must suffer death; this is what justice, as the idea of judicial authority, wills in accordance with universal laws that are grounded a priori. (6:334)

And on the subject of granting clemency:

> Of all the rights of a sovereign, the *right to grant clemency* to a criminal…is the slipperiest one for him to exercise…With regard to crimes of *subjects* against one another it is absolutely not for him to exercise it; for here failure to punish is the greatest wrong against his subjects. (6:337)

In each of these three passages, Kant expresses the strict necessity characteristic of his retributive view. While each described a different aspect of punishment, there is a common element to the retributivism expressed by them: it is of vital importance that punishment be applied to all and only those who are guilty of committing crime. While one might readily draw the conclusion from these quotations that Kant supports a

[2] *Ius talionis* plays a significant role in Kant's views on both the appropriate amount and method of punishment, and I discuss it at length in relation to these two issues in the next chapter. For now, suffice it to say that Kant sees the law of retribution as requiring that we punish those who have done wrong.

retributive justification for punishment, that view is actually not expressed. Even when Kant says that punishment is applied 'because' the subject has committed a crime, there is some ambiguity in his meaning. While it is natural to read this as specifying the justification for the punitive action, it is also possible to read it as picking out who is punishable. According to this reading, Kant is stating that the state can only punish those who have previously committed a crime. Although this is a retributive claim, it is not one that necessarily requires a retributive justification.

It might be tempting to view this distinction as ultimately insignificant. Even though Kant is referring directly to punishing those who have acted wrongly, it is clear that he thinks punishment is itself justified as a direct response to wrongdoing. Indeed, the Kantian picture holds that the underlying retributive justification for punishment is the basis for selecting a retributive liability, and it is possible that he does not recognize the distinction between these two things. This connection, however, is not strictly necessary. It is possible for a theory to support a retributive principle for specifying liability within the broader context of a deterrent justification. It is precisely this balance that will allow for Kantian protective deterrence to preserve many of the statements and much of the spirit of Kant's retributive theory.

What would this coexistence look like? At its most basic, such a retributive liability would describe concrete rules of practice for the executive, judicial, and legislative branches of government; these rules would impose certain retributive constraints on the practice of punishment. Thomas Hill describes this division in his paper "Kant on Wrongdoing, Desert, and Punishment" in the following way:

"deterrence plays a role in the general justification of the practice of punishment, but this is compatible with retributive policies governing judges, juries, and even legislators operating within the framework of the practice."[3] In short, although deterrence provides the reason why the state is authorized to punish, the laws and practices of punishment are not set up solely to achieve the maximum amount of deterrence possible; the state's practices achieve deterrence through means that are constrained by other concerns, specifically ones that appear retributive in character.

If Kantian protective deterrence endorsed a retributive liability, this would entail several significant consequences for the manner in which crime is prosecuted. The judicial system, including both the executive branch's investigation of crime and the judicial branch's trying of cases, should arguably be conducted in the language and according to the guidelines of retributivism.[4] When handing down sentences, for instance, judges would make reference to the crime committed, rather than directly justifying the sentence in terms of the need to preserve the deterrent efficacy of the threat that the law represents. While the law and its corresponding sanction are being designed, the penalty for its violation should be designed with deterrence in mind; once the law has been violated and the penalty must be imposed, however, it is not only possible but potentially desirable to do so in a manner with strong retributive elements.

---

[3] Hill, Thomas E., Jr. "Kant on Wrongdoing, Desert, and Punishment." *Law and Philosophy*, Vol. 18, No. 4 (Jul., 1999), pp. 430.

[4] In the next chapter, I shall argue that this retributive appearance need not – and perhaps *should* not – extend to the actual content of the penalty applied. It would still be possible to use retributive language during a trial and then adopt a rehabilitative framework while the sentence is carried out.

There are two reasons for why a Kantian deterrence theorist might want to endorse policies of this kind. First, the use of a retributive liability is the only way that we can hope to capture any of Kant's explicit retributivism within the framework of a deterrent theory. The passages quoted above – and more like them – could be preserved, rather than jettisoned. Although this marriage will be imperfect in one or two respects, it nevertheless represents the strongest, most Kantian option available. Second, there are significant pragmatic considerations: it might well be the case that utilizing the language of retributivism in a limited way during the practice of punishing might be most effective at deterring crime. This possibility is, to a degree, reminiscent of the so-called 'paradox of hedonism.' As explained by Mill, Sidgwick, and others, if our goal is to attain happiness or pleasure, we must aim at some other good; if we intentionally sought only our own happiness, we would fail to achieve the greatest amount of happiness possible. In the same way, we might think that making punishment's aim of deterrence explicit would actually decrease the amount of deterrent force that our penal practices can generate. Only by linking crimes and punishments in a direct, retributive manner could the state hope to truly convey a properly deterrent message to would-be criminals.

It seems reasonable, then, to suggest that the retributive fixing of liability would substantially support the overall deterrent justification of the institution or practice. This harmony between the deterrent justification and retributive liability of punishment is typical of mixed theories of punishment. A mixed theory of punishment is one that utilizes different, seemingly incompatible rationales at different levels of the theory.

While a classic retributive or deterrent theory is, respectively, retributive or deterrent in all of its various elements, mixed theories might include retributive, deterrent, and rehabilitative elements. Traditionally, Kant is interpreted as adopting an unmixed, retributive theory in which all questions – be they the justification, the liability, the amount, or the method of punishment – are settled by reference to retribution and moral desert. Kantian protective deterrence, on the other hand, will be a highly mixed theory that incorporates retributive and rehabilitative elements into a fundamentally deterrent theory.[5]

There is a different usage of the term 'mixed theory' that is occasionally employed in the literature on punishment. According to this conception, a mixed theory is one that supports multiple justifications for punishment. Jean-Christophe Merle employs this sense of mixed theories in his paper "A Kantian Critique of Kant's Theory of Punishment."[6] A significant part of Merle's paper is given over to criticism of Byrd, Hill, and others who have advocated Kantian deterrence. His criticism is largely founded on the claim that these efforts purport to be mixed theories, but are ultimately

[5] Mixed theories are not without their weaknesses. While they have the advantage of capturing many of the various interests and expectations we have for a system of punishment, they are also prone to greater internal tension. While it might seem unproblematic to say that the justifying aim of deterrence is well-supported by a retributive liability, a deeper investigation threatens to reveal that the rationale behind the liability is either incorrect or misleading. In short, some express doubt that a liability could be adopted that would run counter to the general justifying aim of the practice. For a full treatment of mixed theories and some of their problems, see Kaufman, Whitley R. P. "The Mixed Theory of Punishment." *Honor and Revenge: A Theory of Punishment*. Dordrecht: Springer Publishing, 2013.

[6] Merle, Jean-Christophe. "A Kantian Critique of Kant's Theory of Punishment." *Law and Philosophy*, Vol. 19, No. 3 (May, 2000), pp. 311-338

still employ an exclusively retributive justification. He goes on to advocate for a more truly deterrent position, albeit one that jettisons most of Kant's original views.[7]

Although there is merit to the position he ultimately defends, Merle's criticisms are based on several significant errors. First, he misconstrues the nature of Byrd and the others' views. He lists a number of theses to which retributivist theories are or can be committed, focusing on the following four:

> (a) All criminals and only criminals should be punished.
> (b) The punishment of the criminal constitutes retribution for the crime committed.
> (c) The degree of punishment should be (ordinally, not cardinally) proportionate to the crime, i.e. the scale of punishments must correspond to the scale of crimes. By this I mean that a more serious crime should be punished more severely than a less serious crime, and that two equally serious crimes should result in punishments that are each as severe as the other.
> (d) The degree of punishment must be equal to the crime.[8]

Merle claims that these four theses remain intact under Byrd's 'mixed' theory. On the other hand, he identifies deterrence as also committed to four theses. He lists them as well:

> i) Future crimes are deterred by the punishment of actual criminals. Contrary to the following two theses, this descriptive thesis does not belong to any theory of deterrence considered to be normative.
> ii) Future crimes should be deterred by the punishment of actual criminals.
> iii) Citizens should be punished in such manner as to provide the most efficient deterrence to the commission of future crimes.

---

[7] Merle develops this line of thought further in his book on Kant, Hegel, and Fichte's theories of punishment (Merle, Jean-Christophe. *German Idealism and the Concept of Punishment*. Cambridge: Cambridge University Press, 2009.) While his arguments against mixed theories remain mostly unchanged and, consequently, unconvincing, I will cite his thoughts on Kant's rehabilitative elements in chapter six.

[8] Ibid., p. 316.

> iv) Criminals and only criminals should be punished, and this should be in such manner as to provide the most efficient deterrence to the commission of future crimes.[9]

According to Merle, Byrd's mixed theory accepts the first two, but rejects the third and fourth. Given that she retains all four of the retributive theses, he contends that her position is, in truth, deeply retributive, with only a descriptive veneer of deterrence.

This is an implausible reading of Kantian mixed theories for several reasons. To begin with, a number of Merle's theses are questionable. For instance, the claim that criminals should be punished "in such a manner as to provide the most efficient deterrence" is not strictly entailed by all forms of deterrence. This is a specifically maximizing conception of deterrence, and while that version has a long history, it is not the only possible version.[10] It is possible to support a version of deterrence that aims at achieving only a satisficing degree of deterrent force. Likewise, the claim that proportionality is built into retributivism as a justification is an unsupported leap. Although many retributivists also defend equivalence between crime and punishment, this is not strictly necessary. It is possible to be a retributivist who holds that criminals deserve to suffer punishment, but that this desert makes no reference to suffering something equivalent to the crime.[11]

---

[9] Ibid., p. 317

[10] There are good reasons to be skeptical of the view that a deterrent theory must aim to punish so as to most efficiently generate deterrence. The terms being employed here are exceptionally vague: what makes deterrence 'efficient?' How reliably deterred must the criminal be? How does one balance the deterrence of the subject of punishment and the vicarious deterrence of third parties? I will further address all of the issues in the next chapter, but for now, suffice it to say that Merle's very imprecise sketch of the commitments of deterrence theorists leaves much room for doubt.

[11] One version of this kind of view might approach moral desert in a positivist fashion. We might think, for instance, that violation of the law renders you deserving of punishment – specifically,

In addition to positing objectionable theses, Merle also makes a mistake by assuming that a theory that accommodates more retributive theses than deterrent ones is necessarily retributive in its justification. Even if we accepted that Byrd's theory supports all four retributive theses but only two deterrent theses, this is not a problem for Byrd. The entire point of a mixed theory is the ability to accommodate elements of another approach to punishment. Merle's theses describe different aspects of a theory of punishment: some make reference to the justification, while others refer to the liability, and still others describe the amount or method of punishment. Byrd's theory accommodates the deterrent theses that refer to justification; the fact that her theory does not support the deterrent theses that refer to liability or amount is not a sign that her theory is retributive in its justification.

In short, Merle has constructed his theses in such a way as to be incompatible with the very idea of a mixed theory. If we reject the possibility of mixed theories, then Merle's argument begins to make more sense. If we accept that a theory can pursue retributive interests within a framework justified by deterrence, however, then his criticisms no longer carry any weight.

Merle's critique also goes wrong in misunderstanding the kind of mixed theory that Byrd, Hill, and others are advocating. Merle assumes that either 1) a mixed theory must truly satisfy multiple justifications, or 2) that a theory's justification is no more important that the principles of distribution it employs. If we correct for these mistakes,

---

whatever punishment is legally associated with such a violation. According to this view, we would do wrong by failing to punish or punishing in any other way, even if it more closely resembled the crime committed.

then it becomes substantially easier to conceive of how a mixed theory might be possible. Instead of multiple justifications, all at the same level of the theory, we should pursue a theory that offers a single justification for punishment, but allows other elements of the theory to be structured in different manners. In this case, what I and others propose is to read Kant as most plausibly able to defend a theory that is justified by deterrence, but that allows for retributivism to play an important role in organizing the way in which punishment is distributed. These principles of distribution must still serve the general aim of punishment – given by its justification – but they can do so in ways that are not dictated solely by immediate considerations of this aim.

Putting Merle's interpretive errors to the side, I think there are additional good reasons to avoid this second conception of mixed theories. A mixed theory that truly allows for multiple justifications for punishment is identical to adopting pluralism about the justification of punishment. According to the central logic of pluralism about punishment, we have a number of radically different interests, concerns, and goals that we want an institution of punishment to address or accomplish. Any policy, institution, or act of punishment could conceivably be fully justified if it successfully accomplished one of these various ends. Thus, there is no single reason why punishment is permissible; there are a number of potential reasons, all of which are equally valid.

Pluralism, however, has difficulty resolving cases in which these various interests are at odds with one another. It seems uncontroversial to say that we regularly face cases in which our interests in retribution, deterrence, and rehabilitation each incline us to punish differently. If any one of these three ends is sufficient to justify a

different course of action, then there is no rational basis on which to choose between them. Likewise, we cannot meaningful resolve a dispute over which one ought to be chosen. This is philosophically undesirable, but it has even more serious practical implications. If there is no single justification on which to ultimately fall back, then it becomes impossible to hold the criminal justice system accountable for certain actions. While rank abuse could still be denounced, on what basis could we object to inequalities in treatment, in which some violators are punished retributively and others rehabilitatively? In a truly pluralistic account, any of these incommensurate interests is sufficient to justify punishment, so it is not possible to say that an individual should have been punished according to some other justification.

A genuine, hierarchical mixed theory avoids these difficulties. It enables us to recognize and accommodate the various interests we have with respect to punishment, without giving up on a single justification. The presence of this justification provides guidance and settles matters of dispute. Likewise, it creates standards of evaluation that allow for judgments about the appropriateness of a concrete act or system of punishment. By incorporating a retributive liability into Kantian protective deterrence, many of the concerns that could be raised about a purely deterrent theory of punishment are satisfied.

5.2 Punishing *Only* Those Who Have Done Wrong

I have already discussed some of the advantages that can be reaped by incorporating within Kantian protective deterrence retributive practices of identifying the individuals who should be punished. In addition to the practical advantages, there is great interpretive value in this, as it allows us to consistently preserve much of Kant's original writing on the subject of punishment. In order for this to be the case, however, Kantian protective deterrence must still account for why punishment must be applied 1) to ***only*** those who have done wrong and 2) to ***all*** those who have done wrong. As I will show, incorporating these two constraints into a fully deterrent theory would be difficult; only by endorsing retributive policies that are subordinate to the deterrent justification can the theory accomplish this goal.

Let us begin, then, by exploring how Kant could meet the criterion of ***only*** punishing those who have done wrong. This is a relatively straightforward task. The solution involves Kant's underlying moral philosophy, specifically the second formulation of the categorical imperative. By coupling this conception of the moral law with what Kant says about the possibility of juridical law, we will see that Kant has ample resources to provide a reason for constraining punishment to be applicable to only those who have violated the law.

This constraint also has the advantage of answering Kant's second formulation argument, discussed in the previous chapter. Recall that alongside his flawed, positive argument that grounded the necessity of retributivism in moral desert, Kant also offers a negative argument that I called the second formulation argument. According to this

argument, any justification for punishment other than retributivism would be guilty of using persons as means to achieve some other personal or social goal. If Kantian protective deterrence can show that only the guilty will be punished, however, then the force of this objection is lost.

The concern with punishing only the guilty is one that derives its force from critiques concerning the limitations of the deterrence tradition, and in particular its consequentialist underpinnings. Kant anticipates a debate that would hound the utilitarian approach to punishment in the nineteenth and twentieth centuries. The classical utilitarians uniformly supported punishing for the sake of deterrence, as the suffering of a criminal was itself a wrong that could only be justified if it promoted greater utility. Opponents of the deterrence theorists saw a potential danger, however, and sought to demonstrate the unsavory conclusions of such utilitarian thinking. Their argument runs as follows: if punishment is only justified for the sake of deterrence and the utility that such deterrence generates, then we ought to determine those who are liable for punishment solely on considerations of how to maximize utility. While we will often get the greatest utility by punishing criminals, there is no necessary link. Furthermore, the opponent of deterrence might state, we might be not only permitted, but indeed morally required to punish an innocent if such an act were in the interest of general utility.

The traditional defense adopted by deterrence theorists is an empirical, pragmatic one. While they must admit that there is no principled reason why the innocent are safe from punishment in a manner that is distinct from wrongdoers, they

argue that punishing the innocent is, in fact, an ineffective or otherwise defective means of generating utility. It is difficult to imagine that widespread punishing of the innocent would actually deter crime, as eventually people would begin to realize that they might be punished even if they did not commit crime, thus incentivizing the violation of certain laws.

The defense offered by the utilitarian deterrence theorists is relatively compelling. Although we would need to do empirical research to confirm their claims, it does seem plausible that a deterrence theory would not allow for punishing the innocent under normal circumstances. This prohibition does rest, however, on facts about the world; if some of these facts were to change, then it is possible the argument's validity might also be affected. No doubt a determined philosopher could construct hypothetical situations in which it truly turns out to be a net gain for utility to punish those who have not done wrong.

Fortunately, we need not contemplate such increasingly speculative arguments about the possibility of various hypothetical scenarios. While the consequentialist camp of deterrence thinkers has no principled recourse to protecting the innocent, Kantian protective deterrence is fully capable of explaining why it is *never* acceptable to punish an innocent, even if such an action would produce the greatest utility for society. As we will see, the solution not only allows us to avoid the traditional criticism of deterrence, but it also provides an opportunity for reinterpreting some of Kant's most retributive passages.

In the Doctrine of Right, Kant writes,

> Punishment can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society. It must always be inflicted upon him only *because he has committed a crime*. For a human being can never be treated merely as a means to the purpose of another or be put among the object of rights to things: his innate personality protects him from this, even though he can be condemned to lose his civil personality. He must previously have been found *punishable* before any thought can be given to drawing from his punishment something of use for himself or his fellow citizens. The law of punishment is a categorical imperative, and woe to him who crawls through the windings of eudaemonism in order to discover something that releases the criminal from punishment or even reduces its amount by the advantage it promises. (6:331)

As this quotation demonstrates, a large part of Kant's reason for endorsing retributivism is his concern that any other justification for punishment must resort to using humans merely as means to some end, rather than as ends in themselves. In the last chapter, however, I argued that this way of thinking, although consistent with Kant's moral philosophy, cannot on its own establish why the state would be justified in a deterrent manner. It is a sound argument, but it acts merely as a negative constraint upon whatever theory of punishment we wish to ascribe to Kant.

It is not difficult, however, for Kantian protective deterrence to meet this constraint. Even though punishment is justified by deterrence, the state could never punish one who has not violated the law; Kant is right to think that such a punishment would use an individual as a means to achieving deterrence in a way that is incompatible with the formula of humanity. The reason why such punishment would involve using persons merely as a means, however, is not because wrongdoers are *deserving* of punishment in any way the state can respond to, and the innocent are not; rather, the reason can be found in Kant's writing on law and the limits of what a state can make a law.

According to the Kantian protective deterrence, criminals can be punished for the sake of preserving the deterrent force of the law without running the risk of using their personhood as a mere means. This is due to their ability to consent to the laws that govern a juridical state. Although a thief might not actually consent to property laws and the penalties they prescribe for those who violate them, as a rational being he or she is not only capable, but required to give consent to such laws. In light of this, the thief is obligated to follow the law or suffer the consequences for refusal. Kant argues that although the thief does not, in fact, will the punishment that follows from his or her action, he or she can be said to recognize its necessity as a rational being:

> As a colegislator in dictating *penal law*, I cannot possibly be the same person who, as a subject, is punished in accordance with the law…Consequently, when I draw up a penal law against myself as a criminal, it is pure reason in me (*homo noumenon*), legislating with regard to rights, which subjects me, as someone capable of crime and so as another person (*homo phenomenon*), to the penal law, together with all others in a civil union. (6:335)

While Kant is careful to claim that the permissibility of punishment does not depend upon the criminal's personal judgment, the fact that the rational humanity of the criminal recognizes the need for the punishment gives us all we need. This recognition allows the state to punish while still acting in a manner that is approved by all involved; it would be strange to suggest that the state acts inappropriately when it follows a course of action that the criminal herself must rationally will.

Can criminals rationally will that they be punished for the sake of deterrence? I believe they can. Recall that the explicit purpose of the state is to determine and preserve the conditions of external right. One of the most effective and least intrusive means the state has to accomplish this goal is through threatening the use of legal sanction. What's

more, this punishment must be connected to the activity that it seeks to deter if it is to be effective. Taking all of these considerations together, I posit that citizens, *qua* rational legislators, face no obstacle to consenting to their use to deter crime, provided that they have engaged in illegal action. The institution of punishing is still justified by deterrent interests, but it limits the pool of citizens that it can punish for deterrent reasons to be those that have previously committed crime. By committing crime, these individuals voluntarily place themselves within the group of citizens that can be punished so as to generate deterrent force. While any citizen's unwilling use would indeed violate the formula of humanity, no one is being used as a mere means in this account.

These reasons do not, however, extend to the possibility of punishing the innocent. Kant is clear in *One the Common Saying* that "What a people cannot decree for itself, a legislator cannot decree for a people" (8:304). He elaborates on what conditions must be met for some policy to be that which people cannot decree for itself:

> If a public law is so constituted that a whole people *could not possibly* give its consent to it…it is unjust; but if it is *only possible* that a people could agree to it, it is a duty to consider the law just, even if the people is at present in such a situation or frame of mind that, if consulted about it, it would probably refuse its consent. (8:297)[12]

Any law that calls for punishing the innocent, I contend, would fail this test; the citizens of a state could not possible consent to or will such a law. Given that the innocent have not violated the law, they have not 'volunteered' for use in the deterrent

[12] If we take this limitation seriously and couple it with the state's role in creating a maximally extensive scheme of equal external freedom for each citizen, then it appears that any law that infringes unnecessarily on even one citizen's freedom would be not only a bad law, but an unjust, unwillable law. There is no obvious way around this conclusion. I return to this question in chapter seven.

system. If the state used coercive force against them, it would violate their freedom without any rational basis for believing that this action is meaningfully connected to the state's interests and purposes. Thus, the rational nature of the citizens could not endorse this punishment in the way that it could for the punishment of criminals. To punish an innocent regardless of this would indeed be to use him or her as a means to an end, thus violating the moral law. Any legal policy that violates the moral law, thereby requiring one or more citizens to act immorally, could not be a possible law. We are left with the conclusion that punishing the innocent would violate the formula of humanity, and thus is always impermissible.[13]

This is why even a deterrent interpretation of Kant's theory of punishment can still fulfill the condition of punishing only those who have done wrong. The limitations imposed by the nature of law and the formula of humanity rule out the possibility of punishing the innocent for the sake of deterrence; to do so would be to treat them merely as a means. The guilty, on the other hand, have rationally consented to their punishment in such a way as to allow for their punishment for the sake of deterrence without violating the formula of humanity. In this way, we capture the most important

[13] We might worry about whether such punishments could nevertheless be carried out, provided they are not specified by law. That is, if the objection to punishing the innocent is that no law permitting such punishment could be legislated, then could this concern be sidestepped by the executive's punishing the innocent in extra-legal actions? In one sense, such extra-legal action could not be an accepted part of any theory of punishment, for it differs in several key respects from Kant's definition of 'punishment,' discussed in chapter three. Yet, it is not clear that Kant would characterize these executive actions as impermissible. I must set this issue aside for the time being, but I will return to it in chapter seven, which addresses revolution and punishing rulers who have abused their positions of authority.

elements of Kant's apparent retributivism: no one is punished who has not done wrong, and no one is treated merely as means to achieve the goal of deterrence.[14]

5.3 Punishing *All* Those Who Have Done Wrong

Moving on, now let us consider if and how Kantian protective deterrence can satisfy the condition of punishing **all** of those who have done wrong. Committing to a prohibition against punishing the innocent is by no means a guarantee that the state will also punish all the guilty. Indeed, the argument used to demonstrate the impermissibility of punishing the innocent does not work in this scenario; to allow a wrongdoer to go unpunished does not represent the direct violation of anyone's right to the free exercise of her or his external freedom.

Nor is Kant's own reasoning available to us. In explaining why all wrongdoers must be punished, Kant relies heavily on the concept of moral desert. He describes those who fail to punish as guilty of collaborating in the act of injustice. As we will see below, he holds there to be no exception to this. The absolute moral necessity of state punishment, however, can find no place in Kantian protective deterrence. We must remain committed to preserving the strict divisions that Kant draws between *Recht* and *Tugend*, and this requires some other basis for state punishment than moral desert.

---

[14] A similar argument is developed in *Scheid, Don. E.* "Kant's Retributivism." *Ethics, 93 (1983), pp. 262-282.* Scheid makes the following statement: "It is clear he regards some principle like…the *jus talionis*…as…required to guarantee that the individual within the system of legal punishment is treated with due respect. Now, given these retributivist principles, we may interpret Kant as saying, roughly, that it is permissible to 'use' a person so long as the treatment is consistent with these principles. The individual is being treated with the respect due him as a person, that is, according to what is justly due him - as established by the retributivist principles."

Unlike the question of how to explain punishing *only* the guilty, Kantian protective deterrence cannot give any objective or universally necessary reason for punishing *all* of the guilty. In this respect, it must fall short of Kant's goal. This does not mean, however, that the theory would allow for the guilty to go unpunished as a matter of common practice. Instead, Kantian protective deterrence bases the need for all wrongdoers to be punished on claims regarding empirical regularities.

What facts enable Kantian protective deterrence to prescribe punishing **all** of those who have been found to have committed a crime? If the state's interest is in deterring any and all crimes, it would best achieve this end by showing that no violation of the law will go unpunished. Failing to punish a crime would serve to decrease a criminal's confidence that she or he would be punished as a result of wrongdoing, thus decreasing the disincentives to breaking the law. This is true for both individual and vicarious deterrence: failure to punish a person's criminal action would likely increase the odds that she would do so again, and learning that she has gone unpunished would likely have the same effect on others. Barring extreme circumstances, then, punishing all who have committed a crime is clearly the most sensible way to achieve the deterrent purpose of the institution of punishment.

This is, admittedly, an empirical claim, and it could be shown to be incorrect by facts about the world. Additionally, one might worry that this manner of response would fail when it comes to cases that involve unusual or rare crimes. We might think of particular crimes that are only possible under a rare or unique set of circumstances; there might be no possibility of recidivism or repeated offenses committed by others.

Regardless of the details, the concern that motivates this type of objection is that, due to the exceptional nature of the crime, the individual would never be inclined or even able to repeat her or his offense; furthermore, it is even possible that widespread, public knowledge of the unpunished crime could not produce copycat crimes, as no other citizen would be in a position to repeat the offense. In such cases, it might not appear that punishing actually deters anyone, and thus perpetrators of exceptional crimes should go unpunished.

This way of thinking is ultimately too simplistic to pose a real challenge to any reasonable deterrence theory, let alone our Kantian variety. To illustrate why this is the case, I think it will be useful to refer back to the 'preservation of the state' arguments developed by Byrd and Ripstein. While I argued in the previous chapter that their related approaches were too indirect to justify a robust Kantian deterrence theory of punishment, I think that they do a nice job – Ripstein especially – of highlighting the flaws in the 'exceptional crime' argument I presented above.

Recall that Ripstein's interpretation of Kant's theory of punishment justifies the state's use of force as necessary for securing and preserving the authority of the law. The law, arising from the unified wills of all citizens, cannot allow criminals to unilaterally exempt themselves from the rules that bind all. As he puts it, "The threat of punishment is thus the announcement that public law will remain supreme."[15] Regardless of the shortcomings of this as the sole justification for punishment, Ripstein is right to suggest

[15] Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge: Harvard University Press, 2009. Pp. 302.

that the law – and indeed the possibility of a juridical state – is threatened by individuals seeking to exempt themselves from its authority. No state can achieve its purpose in securing and guaranteeing external freedom if its members constantly seek ways to undermine the rule of law.

It is precisely this danger that rules out the possibility of exceptional crimes going unpunished by a deterrent theory. In the case of widespread public knowledge of the criminal's escape from punishment, the respect the public holds for the law and its authority will be greatly weakened. Even if they cannot commit a similar crime, they learn the lesson that the state is powerless to enforce its threats in special circumstances and might begin seeking out such situations in their own lives. In the case where the public is not aware of a crime's having gone unpunished, the criminal will still be aware. While she or he might never have a reason or opportunity to repeat the offense, escaping penalty will still have the same detrimental effect of her or his respect for the law. Even in the case of exceptional crimes, then, punishment still serves a deterrent effect by reinforcing the supremacy of the law and the unfeasibility of exempting oneself from it.

Nevertheless, a defender of retributivism might point out that it is possible to create a hypothetical counterexample that shows the deterrence line of argumentation to be lacking; indeed, Kant has already done just this in his notorious 'dissolving state' example. In perhaps his best known statement on punishment, Kant writes,

> Even if a civil society were to be dissolved by the consent of all its members…the last murderer remaining in prison would first have to be executed, so that each has done to him what his deeds deserve and blood guilt does not cling to the people for not having insisted upon this punishment; for otherwise the people can be regarded as collaborators in this public violation of justice. (6:333)

This passage reveals two things. First, it demonstrates the true depth of Kant's commitment to a retributive liability. Those who have done wrong – murdered, in this case – must be punished, regardless of even the most extreme of social, political, and legal circumstances. Second, it exposes the limits of Kantian protective deterrence's ability to accommodate Kant's retributive claims. There is no way to account for the necessity of punishment in this kind of scenario under a deterrent theory; after all, there is no future rightful condition for the state to preserve. Its interest in protecting the freedom of its citizens is ending, and the punishment of past violators does not serve that interest any more. As such, this passage cannot be accommodated within Kantian protective deterrence, and we are left with no option but to reject it as deeply incompatible with the best, most consistent elements of Kant's practical philosophy.

Despite these limitations, I have three responses to the dissolving state example. First, it should be noted that the dissolving state example is not, in itself, an argument. It is a claim meant to demonstrate the full force of Kant's commitment to retributivism. It is conceivably meant to rule out the very kind of deterrent reinterpretation that I have proposed. However, we should not abandon our deterrent project in light of this claim; if the two are irreconcilably in conflict, then we should abandon the dissolving state example, rather than Kantian protective deterrence, as the latter enjoys strong support from Kant's more foundational moral and political philosophy.

Even if we ignore its lack of reasoned support, there is much that remains unclear about the dissolving state example, and this imprecision leaves open some very

difficult questions. While Kant makes it clear that murders would need to be executed before the state fully disbands, he says nothing about those found guilty of other crimes. If he truly supports a full-fledged retributivist theory, then at a purely qualitative level, the strength of the desert should be equal regardless of the crime committed; either one deserves to be punished or one does not. Although the nature of the crime might matter in determining the amount and nature of the punishment, all offenses should result in the same kind of deservingness for punishment. If this is true, then there is no reason why Kant should single out murderers in the dissolving state example. All convicted criminals should be punished in the appropriate manner before the members of the state can go their separate ways.

If this is so, however, then it is not clear how Kant would handle the fulfillment of prison sentences. If a criminal is sentenced to ten years imprisonment, must the state wait ten years to dissolve? The law of retribution would forbid us from altering his or her punishment in light of the circumstances. This conclusion seems absurd, and it becomes even more so when we consider that the state would still be obligated to punish those who commit wrongs while waiting for the state's dissolution to occur. These punishments could in turn prolong the waiting period, thus making the final dissolution of the state a practical impossibility. While Kant's example might seem initially warranted within a retributive theory, it quickly becomes self-defeating when we broaden it to crimes other than murder.

Third, there is another reason to think that this is not, in truth, a good counterexample for Kant. At the risk of stating the obvious, the dissolving state example

describes a situation in which the members of a state have unanimously elected to dissolve the state and go their separate ways, dispersing throughout the world. This latter condition is important; Kant could not allow for the dissolution of the state if its members had no plans to leave their present location and each other's company. The proximity of other humans and our inability to live without interacting with them are some of the fundamental contingent facts about human anthropology that make living within a state a necessity, indeed one that can be forced upon those who refuse.

Given contemporary facts about the size of the human population, the extent of human civilization across the surface of the earth, and the size of modern countries, the situation that Kant describes in the dissolving state example might no longer be possible. The possibility of acquiring the consent of *every* member of a state alone seems remote enough to render it virtually inconceivable. Even if such consensus was successfully reached, however, it is not clear that the state's dissolution would still be permissible; the former citizens of the state might not be able to disperse in the requisite way. If even a small group of them remained behind, these individuals would be required to share a civil state with one another. While it might be possible for them to reform a new state in the absence of the others who have departed, this move seems far too close to secession or change of government, options that Kant clearly forbids (8:298-302).

My second and third points are, admittedly, practical objections to a theoretical hypothetical. The fact that it is not realizable does not change the nature of what would be required if it were. Perhaps, though, this should serve as a *reductio* of sorts on the extremeness of Kant's own view; it might be absurd to think that any functional theory

of punishment could account for and explain a situation as bizarre and unrealistic as this. As I have shown, the dissolving state example is deeply unclear, beset by problems, and perhaps prohibited by Kant's more fundamental political philosophy. Coupled with the weakness of Kant's arguments for retributivism, we ought not concern ourselves that this thought experiment poses a serious risk to the Kantian protective deterrence.

Might it be possible to construct a different kind of case, one that articulates Kant's retributivism while avoiding some of the difficulties that the dissolving state example faces? Quite possibly. In the face of such examples, Kantian protective deterrence can ultimately offer nothing more than the assurances that, in all ordinary cases, all of those who have committed a crime must be punished, if for no other reason than a failure to do so would diminish the efficacy of the law's deterrent force. As strong as this likelihood is, it still remains a contingent matter; there is no principled reason why the state must punish all of those who have done wrong. As such, some of Kant's commitments will be forever incompatible with the view I am developing.

From a purely interpretive standpoint, this is perhaps a slightly disappointing result. While previously I have had some success reinterpreting Kant's more retributive sounding passages, in this case there will be some that simply must be rejected as inconsistent with the deterrence theory I am advocating. It would clearly be preferable to incorporate all of his statements on punishment in one consistent unity, but this simply is not possible. The principle of charity would plausibly lead us to preserve as many passages as possible within a coherent unity, rather than aim for complete retention beset by contradiction.

Perhaps, however, we might find a silver-lining of sorts in this small inconsistency. The most prominent passage that will need to be jettisoned is also one of Kant's most notorious: the dissolving state example has long sat uneasily with many readers of Kant. There is ultimately no way to align this claim with a deterrence theory, but given the lengths to which Kant scholars have gone in order to explain away his way of thinking in this thought experiment, perhaps it is not a claim that should be accommodated in a Kantian theory. Rather, perhaps Kant's position will be strengthened by abandoning this difficult and controversial example and the philosophical commitments that are meant to support it.

Reading his retributivism as a specification of liability for punishment not only avoids these difficulties, it also enables the construction of a flexible, robust mixed theory of punishment. Many of Kant's concerns about the dangers of a punitive system justified by and organized solely around an interest in generating deterrence are well founded. A theory of punishment that incorporates no significant 'retributive' constraints runs a serious risk of clashing with other, more fundamental elements of Kant's practical philosophy. Although a purely deterrent approach to punishment might, as a practical matter, prescribe many of the same policies and practices that Kant views as essential – for example, punishing only the guilty – it would do so for reasons that Kant should ultimately reject as limited or misguided. According to Kantian protective deterrence, an interest in deterring crime might be what justifies that state's use of punishment, but the application of this power should be meaningfully

constrained by the overriding concern for treating all citizens as ends in themselves. This interest can be best guaranteed by incorporating a retributive liability for punishment.

6 The Methods and Amount of Punishment: *Ius Talionis* vs. Rehabilitation

In addition to specifying why punishment is justified and who ought to be punished, a complete theory of punishment needs to identify the principles according to which punishment will actually be carried out. Knowing why we punish and whom we ought to punish is largely irrelevant without further specification of the form that punishment will take, or how much of it is warranted. It is in answering these concrete questions that a theory of punishment has its most palpable, practical effects. While these answer might be guided by considerations about the justification of the institution as a whole, the immediate experience of those who are punished by the state will be shaped primarily by the form and amount of penalty imposed. As such, these elements of a theory are of the greatest importance to the lives of the actual citizens who live within the state.

This chapter considers together the questions of what methods and amount of punishment are most consistent with the theory of Kantian protective deterrence. There are two reasons for addressing these two elements of the theory together. First, it is difficult to provide any strong specification of one without involving the other. Any answer as to the appropriate amount of punishment seems to require a specific conception of the form that this punishment will take, and *vice versa*. While it is possible to speak about equivalence between crime and punishment in only one respect or the

other in some very specific cases, most instances of punishment cannot be fully described without discussing the method and amount together.

The second reason for addressing the method and amount of punishment jointly lies in Kant's own position on these questions. Kant famously argues in favor of the strict law of retribution, *ius talionis*.[1] According to the principle of *ius talionis*, punishment ought to resemble the original crime as much as possible, both in form and in quantity. This sense of strict proportionality finds expression in the common saying "an eye for an eye." For Kant, *ius talionis* serves as the basis for selecting both the method by which crime is punished and the amount of punishment that is applied to the criminal. For both principled and pragmatic reasons, Kant thinks that the purest, most just form of punishment is to do to the criminal what he or she has done to someone else. He writes,

> But what kind and what amount of punishment is it that public justice makes its principle and measure? None other than the principle of equality (in the position of the needle on the scale of justice), to incline no more to one side than to the other. Accordingly, whatever undeserved evil you inflict upon another within the people that you inflict upon yourself. If you insult him, you insult yourself; if you steal from him, you steal from yourself. But only the *law of retribution* (*ius talionis*) – it being understood, of course, that this is applied by a court (not by your private judgment) – can specify definitely the quality and the quantity of punishment; all other principles are fluctuating and unsuited for a sentence of pure and strict justice because extraneous considerations are mixed into them. (6:332)[2]

---

[1] Traditionally, the law of retribution has been referred to by the Latin name of *lex talionis*. Kant's decision to use *ius* instead of *lex* is not one that he ever directly explains or addresses, but the different shades of meaning between the Latin words *ius* and *lex* give some hints as to his reasons. While *lex* means law in a very literal sense – a mandatory edict promulgated by one with legally binding authority – *ius* has more abstract connotations. It refers broadly to the concept of law or legislative obligation, to the system of law in general, and to traditionally recognized rights and duties individuals have under the rule of law. Furthermore, *ius* is sometimes translated as 'right;' given Kant's usage of *recht* and his focus here, his reliance on *ius* rather than *lex* is unsurprising.

[2] Kant, Immanuel. *Metaphysics of Morals*. Trans. by Mary J. Gregor. Cambridge: Cambridge University Press, 1996. All internal citations refer to the standard Prussian Academy edition, volume and pages.

Analyzing the strict proportionality between crimes and punishments is simpler in some cases than others. Kant often uses the example of execution, as it is straightforward in both questions of method and amount: when a person murders another, execution resembles the crime in both respects. As we will see, however, this kind of strict proportionality is difficult or impossible to maintain in many other kinds of cases. Kant himself recognized some of these limitations, but he has no answer for either the full scope or depth of the issues caused by literal adherence to *ius talionis*.

I argue that Kant ought to abandon this literal understanding of the law of retribution. The obstacles to *ius talionis* are too great, even for a theory with a retributive justification. Trying to incorporate it into Kantian protective deterrence would prove impossible. Instead, I argue for the use of a metaphorical proportionality between crime and punishment that could serve as the basis for selection the method and amount of punishment. Unlike prior attempts at metaphorical proportionality, however, I contend that Kant ought to rely on rehabilitative methods. Not only do such methods of punishing provide a version of proportionality that avoids the difficulties facing *ius talionis*, they find interesting support from Kant's lectures and from the *Critique of Practical Reason*. By specifying rehabilitative methods of punishing – and fixing the amount in a broadly deterrent manner – Kantian protective deterrence is able to provide a more complete, nuanced account of the way in which individual persons experience the institution of punishment.

6.1 The Methods of Punishment

Many of Kant's statements on the topic of what methods should be used to punish criminals take the form of prohibitions against acts that would violate moral obligations to our fellow beings. These negative claims do not directly define the acceptable methods of punishing, but they do help to establish certain constraints. No matter the nature of their crimes, criminals retain their humanity, and as such they can never be treated as objects or as mere means to some other end. It is on this basis that Kant rules out the use of torture, mutilation, or other similar forms of punishment (27:552-553). Indeed, direct corporal punishment on a whole is ruled out on the basis that it treats the body of a person as an object.[3]

When it comes to the kinds of punishments that are acceptable, however, Kant provides less specific instruction. The guidance he does give is in the form of a general adherence to the principle of strict proportionality, as exemplified in retribution. According to this specific conception of *ius talionis*, the only way to proportionally answer a crime is to make the criminal experience an event as close as possible to the criminal action. If the criminal steals property, she should lose her property. If she

[3] Whether Kant himself defended the view that all corporal punishments are unacceptable is a bit difficult to determine. On the one hand, he clearly supports the prohibition of punishment that would mutilate. On the other hand, he supports the use of very difficult labor as a penalty for some crimes. The line between this and corporal punishment is not easy to determine. Certainly, in both cases the body of the punished is used to accomplish some other end – namely, suffering. Kant seems to be drawing the line at any violence against the body of the criminal that would maim, incapacitate, or otherwise permanently harm him. Whether there is a good basis for drawing such a distinction, however, remains dubious.

murders, then she should die. He goes so far as to claim "All substituted means of punishment are lacking in proportion, and degenerate into mere arbitrariness" (27:555).

There are several reasons behind Kant's support for this literal interpretation of *ius talionis*. First, he maintains that this kind of strict proportionality is demanded by moral desert. A criminal simply deserves to experience the same thing as what she made another experience. Second, punishing via methods that resemble the crime is, in Kant's view, a good way of ensuring that the amount of punishment also remains proportional. We will return to this issue below, but for now, suffice it to say that Kant thinks it will be easier to punish in the appropriate amount if we punish crimes in the same manner as they were committed. This is the sense in which he is speaking when he claims that any other method degenerates into "mere arbitrariness."

From this, we can conclude that Kant imagines the appropriate method of punishment to be that which most closely resembles the crime committed, provided that this does not violate the moral dignity of the criminal. Thus, we know that murderers should be executed, but rapists may not be raped, and those guilty of bestiality should only be exiled (6:363). While perhaps not a perfect guide, this constrained *ius talionis* is at least clear in some instances.

Unfortunately, this conception of *ius talionis* has difficulties beyond what Kant recognizes. First, there are a number of crimes for which there is no clearly proportional penalty. It is easy to say that if I steal, I should have my property taken away. This becomes difficult, however, if I do not have any property. In punishing a property-less

thief or arson, then, the state would need to employ some other method. This threatens the literal understanding of *ius talionis*.

Second, there are a wide range of cases in which it is not even clear what penalty could possibly resemble the crime. How ought I to be punished, for instance, if I threaten others with harm or property damage, but do not cause actual harm? Alternatively, what if I am guilty of jaywalking or trespassing? There are a wide range of legal offenses that do not seem to allow for any kind of reciprocal, proportionate punishment. Any punishment that the criminal could be made to experience will necessarily differ from the crime in significant respects. *Ius talionis* does not seem able to provide a clear answer about how to punish in these cases.

Third, within the set of crimes that do not seem to allow for a correlative punishment, there is a special group that deserves some attention. Consider cases in which one lies on an official form, cheats on his taxes, or smuggles illegal goods into the country. In each case, the violation at the heart of these crimes does not appear to be against any individual's right to bodily integrity or ideal ownership over property, but rather against the state itself. While it is possible to explain the way in which such actions do infringe upon the particular freedoms of our fellow citizens – e.g., any damages against the state must be compensated by the rest of the citizen – to attempt to punish with strict resemblance in these cases seems a doomed enterprise. In addition, even a non-damaging criminal actions of this sort still threatens the authority of the state; as I argued in chapter four, while this is a poor justification for all legal sanctions,

deterring such threats is a sufficient reason to punish in this particular kind of case. Nevertheless, these crimes do no admit of any obvious parallel method of punishing.

On this particular worry, Kant has a partial answer. When discussing treason, Kant prescribes execution as the appropriate punishment (6:320). Although the state may not truly face a genuine threat to its continued existence, he characterizes treason as an action that aims at the 'death' of the state, and thus death is an appropriate response. In this way, he preserves the resemblance of crime and punishment. Even as ready as Kant is to suggest execution, it does not seem as though every crime that harms the state could be handled in this way. Trespassing on government property, for instance, presumably should not be punished by death.

Fourth, given that Kant's version of *ius talionis* allows for such substitutions in certain instances, he has additional worries. When moral constraints or the impossibility of replicating the action in a penal setting renders substitution necessary, Kant faces a challenge in explaining why a particular substitute activity is more similar to the original criminal action than some other. Consider the case of my lying to some public official. Should I be fined, imprisoned, or required to perform some difficult labor? If *ius talionis* is our guide, then it seems as though there is no ready criterion to help us choose between these alternatives. None of them resembles my original action, and it is not clear that any of them resembles my original action any more than any of the others.

In light of these various concerns, it becomes clear that Kant so often relies on the example of execution because it is one of the few scenarios in which it is clear what penalty would most resemble the crime. In almost all cases, the method of punishment

looks very little like the original criminal act. If Kant is right that any substituted method lacks proportionality, then it would appear that proportionality is a lost cause. If we wish to preserve some semblance of proportionality, we must turn to a less strictly literal version than *ius talionis*.

Above, I demonstrated that Kant's reasons for supporting *ius talionis* are 1) his understanding of the demands of moral desert and 2) his belief that resemblance in the method of punishment will help to ensure that the amount of punishment remains proportional. As I argued in the fourth chapter of this dissertation, this use of moral desert does not fit within the fundamental elements of Kant's practical philosophy. According to Kantian protective deterrence, the state is not punishing in light of moral desert, but rather to accomplish deterrent goals, in service of its role in protecting its citizens' free exercise of their external freedom. Without moral desert, the only reason for supporting *ius talionis* as the principle for determining the method of punishing is the belief that it helps to ensure the proportionality of the amount of punishment. Below, I will argue that this function of *ius talionis* is unnecessary and unhelpful – Kantian protective deterrence has better options available.

6.2 The Amount of Punishment

If *ius talionis* has difficulty explaining what method of punishment ought to be used against a wrongdoer, it faces even greater difficulties in explaining the amount. In addition to specifying the form that punishment will take, a comprehensive theory will also need to explain how much punishment is appropriate. If the state determines that

imprisonment is the correct punishment for some offense, there remains the question of for how long the offender should be imprisoned. Some exceptions aside,[4] any instance of punishing must include a specification of amount. In Kant's traditional, retributive picture, this specification accords with the principle of *ius talionis*.

When applied to this issue, *ius talionis* tell us that the quantity of the punishment ought to be precisely equal to the quantity of the crime. The punishment should be neither more nor less severe than the crime itself. This kind of literal equivalence between crimes and punishments is relatively simple to work out for more basic violations. If I cause damage to another's property, I must pay the victim an amount equal to the cost of the damage. The balance between crime and punishment is also straightforward in the case of murder. If I kill another, I in turn must be executed by the state. Although I will deal with the permissibility of execution more fully below, this is perhaps one of the clearest examples of equivalence in quantity between crime and punishment; both the victim and the murderer lose a life.

Unfortunately, not all crimes and punishments exemplify the clarity and ease of the law of retribution as well as murder. The picture of equivalence becomes fuzzier when dealing with many other kinds of crimes that are regularly committed. There are many types of criminal activity whose wrongfulness, damage, or harm are difficult to establish in quantitative terms. This applies both to so-called 'victimless' crimes and those that affect persons in some non-financial way. How are we to properly value the

[4] Execution is the most obvious exception to this necessity. If a criminal is condemned to death for a crime, this obviously entails the amount of punishment as well. Barring failures in capital procedures, a person cannot be more executed.

wrongfulness of an assault or, even more difficult, the threat of assault? Such things seem impossible to quantify.

This difficulty sheds some light on why Kant is so committed to preserving equivalence between the crime and the method of punishing. Without it, it is not clear how we could ever make the criminal suffer an equal amount of harm. This is further complicated by an example that he frequently uses when discussing punishment. Using Scottish rebels as his example, Kant argues in favor of execution, on the grounds that this treats the honorable rebels fairly and the dishonorable rebels harshly (on the assumption that the latter would prefer life in prison, and the former would rather die) (6:334-335; 27:555). Here, Kant seems concerned with proportionality not between the crime and the amount of punishment, but between the inner character of the criminal and the amount of punishment. Trying to accommodate this element as well renders the possibility of achieving some kind of proportionality virtually impossible. Even setting aside the epistemic difficulties in the state assessing the inner character of an individual accused of criminal wrongdoing, the need to proportionately match the near infinite possible gradations in a person's character makes such a policy infeasible.

A further problem for *ius talionis* stems from its direct proportionality. Many of Kant's predecessors in the early modern period explicitly rejected relying on *ius talionis* due to its inability to actually punish wrongdoers. Imagine a simple case: I steal $100 from my neighbor. According to one literal interpretation of the law of retribution, my punishment ought to be $100. Clearly, this amounts to no punishment at all, regardless of whether one is an advocate of retributivism or deterrence. This seems to be a rather

uncharitable interpretation of *ius talionis*, however, and one that could be easily addressed. It would be perfectly reasonable for a proponent of *ius talionis* to argue in favor of both returning the original money, as well as paying an additional fine of $100. In this way, the criminal suffers the very thing she tried to inflict upon another.

Even with this clarification, however, we might think that this kind of proportionality would fail to truly deter crime. When weighing the potential gains and losses, this low-stakes form of punishing might not convince some to abstain from committing impermissible acts. It might well be the case that many individuals would consider such crimes worth the risk. This also raises additional concerns about the possibility of the wealthy 'buying' the right to commit illegal actions – a concern that Kant takes seriously and attempts to prevent in the case of slander (6:332). The failure to deter crime effectively might be of no consequence to committed retributivists or adherents of *ius talionis*, but it certainly represents an additional reason for Kantian protective deterrence to be skeptical of the viability of *ius talionis*.

There are further problems for this strict proportionality that arise from the limits of punishing. If stealing $200 is punished twice as harshly as stealing $100, then we should expect to see this kind of increase in the amount of punishment for more serious crimes of all kinds. There are limitations to this, however. If a crime is to be punished with life in prison, then proportionality demands punishing a worse crime with a harsher sentence. Yet, adding more time to the prisoner's sentence does not make the penalty worse. Likewise, if anyone who commits a murder must be executed, then it seems impossible to identify a proportional sanction for someone who commits two or

more murders. At some point, it seems as though we would reach an upper threshold – and, most likely, a lower one as well.

This leads us to the greatest difficulty for using a literal interpretation of the law of retribution as a means of determining the amount of punishment that is justified. Put simply, there are many crimes whose damage cannot be directly counted, weighed, or measured in a way that could be equivalent with a penalty. In some cases, this is due to the crime's failure to result in any quantifiable harm, damage, or loss of external freedom for another citizen. If a person drives while intoxicated but does not get into an accident or fires a weapon in a public building but fails to hit anyone, it is not clear what the cost of her crime is. While such actions ought to be prohibited in light of the danger they pose or the rights they violate, the sanction attached to such laws remains mysterious. In some cases, it might be enough to punish based on the likely outcome or what was intended, rather than the criminal's success; in this way, attempted murder, although it wrongs no one, would still be punished as seriously as actual murder. While this approach still faces difficulties – for instance, should we assume that my driving intoxicated will kill one or more people, or simply injure them? – it at least attempts to provide a means of quantifying hard cases.

But what of instances in which, although I intentionally break the law, I intend to do no harm? Imagine that while taking a shortcut to save myself time, I trespass on another's property, but do no damage or cost the owner any expenses. Although I have violated the law, there is no clear way to make sense of what my punishment ought to be. Any number of public crimes could potentially be of this sort; while it is clear that

the rules exist for good reasons and their violation must be punished, the actual violations in question are not quantifiable in a way that would allow for an equivalent penalty.

Proportionality between crime and punishment is important. This insight has been shared by the majority of philosophers and theorists who have ever written on the subject. Whether this interest is supported by deterrent claims – disproportionate punishment is ultimately ineffective at deterring crime – or retributive claims about moral desert, it is clearly one that Kantian protective deterrence ought to try to accommodate. *Ius talionis*, however, is not a feasible way maintain this proportionality.

6.3 Alternatives to *Ius Talionis*

Adherence to a strict interpretation of *ius talionis* is practically untenable. Simply put, there are too many implications of such a literal approach to proportionality that Kant could not accept within his theory. In an effort to preserve the proportionality between crimes and punishments, Kant interpreters have suggested employing non-literal understandings of proportionality. These metaphorical approaches to proportionality come in several varieties, both retributive and deterrent. While they are improvements to the strict law of retribution, they still face many obstacles.

The first alternative to consider is a metaphorical understanding of *ius talionis* that arguably remains closest to Kant's original literalism. Advocated by Jeffrie Murphy, this retributive interpretation holds that Kant should have argued that the sense of

proportionality that *ius talionis* requires is equivalent ordinal position, rather than trying to achieve some kind of cardinal equivalence. Murphy writes,

> P is proportional to C if and only if P, ranked on a scale of punishments from least to most severe, stands on the scale of punishments at the same point that C, ranked on a scale of crimes from least to most serious, stands on the scale of crimes.[5]

The basic argument here is straightforward: imagine that we develop a scale with ten crimes and ten punishments, each ranked from most severe to least severe. After we have these two scales, we can establish proportionality by linking the most severe crime to the most severe punishment, and so on down the list.

While this approach has the virtue of incorporating Kant's concern for proportionality into a more workable system, I think that the ordinal ranking of crimes and punishments still faces serious problems. To begin with, we would need a strict ordering of all crimes and punishments. This would require a state not only to list every single crime and punishment possible, but it would need to establish a hierarchy for each list. The prospects of accomplishing this task seems inconceivable, especially in the case of crimes.

For punishments, there is one way to avoid these difficulties and construct a system of penalties that could be ordinally ranked. This solution would require us to abandon systems that rely on a variety of different kinds of punishment (e.g., imprisonment vs. fines), instituting instead a system that exclusively utilizes a single form or method of punishment. For example, we could create a criminal justice system

---

[5] Murphy, Jeffrie G. "Does Kant Have a Theory of Punishment?" *Columbia Law Review*, Vol. 87, No. 3 (Apr. 1987), p. 530.

that punished through imprisonment or fines alone. In such a case, it would be relatively easy to ordinally rank penalties, for each additional day spent incarcerated or each additional dollar fined would represent a step up on the scale of punishment.

This would, of course, run counter to the concern that Kant expresses for proportionality in method of punishment as well. Murphy's proposal focuses exclusively on proportionality in the amount of punishment, without reference to preserving the resemblance of crimes and punishments. So long as we want a Kantian theory to capture proportionality in both method and amount, this kind of solution would be unavailable. Murphy's proposal, then, seems to face a dilemma: as long as there are different methods of punishing, weighing them against one another seems impossible; using only a single method of punishing avoids this problem, but it seems to sacrifice an important feature of the retributivism that Murphy is defending.

If creating an ordinal ranking of punishments is difficult, achieving a similar ordinal ranking of crimes is even less feasible. To do so, we would need to be capable of comparing any two crimes and determining which is worse. This would likely require the very kind of quantification of crime that proved difficult for the literal *ius talionis* in the first instance. One might try to answer this challenge by proposing some kind of consensus of Millian 'competent judges.' Clearly, it is uncontroversial to say that assaulting another person is worse than stealing an object of little value from another. There is a limit to this consensus, though, and to try to determine whether minor physical harm or serious emotional harm as a result of harassment, intimidation, and threats is worse is to engage in a hopeless enterprise. The very possibility of competent

judges seems unattainable, as these individuals would have to have experienced every possible crime.

Even if we could solve for this problem, though, the actual equivalence between the ordinal ranking of crimes and the ordinal ranking of penalties in Murphy's account does not seem guaranteed to resemble anything like what Kant suggests. To illustrate this point, imagine we live in a society in which there is a maximum fine for theft or property damage. We. It is quite likely that one could do damage in excess of this amount, meaning that this crime would be met by a penalty that is of a lesser quantitative value. For stealing $100 from my neighbor, I might only be fined $25, because this crime and punishment occupy the same position on our ordinal ranking. If this is the case, we have preserved the spirit of *ius talionis* in a way that seems to directly contradict the letter of *ius talionis*.

Murphy gives no defense of this version of proportionality; perhaps he merely sees it as the best of a group of bad options available to Kant. While he is right to suggest that we ought to preserve some kind of connection between the worst crimes and the worst penalties (a feature that Byrd and Ripstein's theories cannot fully accommodate, as discussed in chapter four), the ordinal ranking system is simply too impractical to be a working alternative. It has the same fatal flaw that undermines the literal law of retribution: the damage or harm of crimes cannot always be satisfactorily quantified.

Rather than trying to preserve the spirit of *ius talionis* by proposing an ordinal ranking of crimes and punishments, a second alternative tries modifying it to take deterrence into account more directly. Altered in this way, the appropriate amount of

punishment would be determined by the amount that is necessary to deter crime. In this scenario, the proportionality is between a specific crime and the amount of penalty that is required to deter it. This alternative avoids some of the problems associated with Kant's commitment to proportionality as necessitated by moral desert, and it seems as though it would be a good fit with Kantian protective deterrence. Unfortunately, deterrence alone cannot meaningfully preserve the kind of proportionality that we are seeking. As we will see, the shortcomings of a purely deterrent approach to proportionality can only be overcome by the introduction of some rehabilitative elements.

The first concern we might have regarding this deterrence view is that the proportionality might very well not track our expectations. According to this view, the proportionality between a crime and its punishment is not an equivalence between the harms of the two. Rather, the equivalence in question is between the punishment and the amount of coercive force it would take to discourage crime. In other words, harsher penalties would not be applied to worse crimes, but rather to those crimes that are harder to deter. Indeed, it seems at least possible that the state would begin punishing tempting, minor crimes – like jaywalking or speeding – more seriously than it punishes serious crimes that individuals are less likely to commit, such as murder. While this result is by no means contradictory or obviously incorrect, it does seem to stray dramatically from both the Kantian and traditional understanding of proportionality.

In addition, there are further obstacles specific to fixing the amount of punishment in a strictly deterrent way. First, it is not clear whether the deterrence we

seek is primarily individual or collective. Any deterrence theory is going to support achieving both kinds of deterrence, but which ought to be prioritized in cases of conflict? Suppose we discover that it takes less of a penalty to discourage a criminal from repeating her or his crime than it does to vicariously deter other potential violators. Indeed, it might even be the case that a penalty harsh enough to effectively achieve vicarious deterrence might even increase the likelihood of individual recidivism, as the individual's future options are diminished. In this case, which amount ought we to prefer?

If we select the harsher penalty in favor of gaining the widest possible deterrence, we still have difficulties to consider. To start, what percentage of the population needs to be deterred by the punishment? Is a penalty that deters 50% of the population sufficient? Must it deter 90%? We might ask the same question about individuals; what likelihood of recidivism are we willing to accept? In addition to these questions, we can say with near certainty that different individuals are likely to have wildly different responses to threats. As Hill writes,[6] imagine a situation in which there is some small minority of individuals who are especially difficult to deter, either vicariously or directly. How much punishment should we be willing to inflict to secure their deterrence?

Finally, there is an additional worry about proportional deterrence that is unique to Kant. If we return to the passage in which Kant first lays out his adherence to *ius*

[6] Hill, Thomas E. "Treating Criminals as Ends in Themselves." *Jahrbuch fuer Recht und Ethik*, Vol. 11 (2003), pp. 30-31.

talionis, we can glimpse part of his motivation for selecting a strict law of retribution as his guiding principle. He writes, "All other principles are fluctuating and unsuited for a sentence of pure and strict justice because extraneous considerations are mixed into them," (6:332). It seems possible, then, that at least part of his concern is that any other means of establishing practical principles of punishment necessarily involve considerations about particular cases, considerations that, due to their variable nature, fluctuate in impermissible ways. Laws and their correlative sanctions, by Kant's own definition, must be universally applicable, meaning that we should not accept a theory that allows for the methods and amount of punishment to vary on a case by case basis.[7] Whatever our alternative to *ius talionis*, it must be capable of providing a stable answer to the question of the appropriate amount of punishment; in this way, much of Kant's reason for objecting would be negated.

Using deterrence to determine the amount of punishment that is appropriate might solve for some problems, but it does not, on its own, guarantee this kind of stability. As we have seen, if our aim is to deter individuals, then in some cases it might be necessary to sentence different criminals to different amounts of punishment. In order for deterrent proportionality to be a full alternative to *ius talionis*, it needs an additional constraint.

[7] There are a few forms of flexibility that might potentially be consistent with the Kantian understanding of law. Flexibility in judicial sentencing, for instance, could conceivably be construed as permissible. In this case, however, the rationale would have to be that the law establishes a range of appropriate punishments, to be adjusted based on specific factors. In truth, the judge would not have flexibility in the sense that we use the term in the U.S. legal system; rather, she or he would still be required to issue determinate punishments in light of particular details.

6.4 The Formula of Humanity and Rehabilitation

I contend that the constraints needed to provide stability and a more traditional conception of proportionality to strict deterrence can be achieved through the introduction of rehabilitative elements into Kant's theory. Although Kant's strict anti-paternalism seems dramatically opposed to rehabilitative efforts, these two concepts need not be working at cross purposes. Given the punishment is already the violation of a citizen's external freedom, against her or his phenomenal will, for purposes the state deems necessary, the very institution of punishment represents one half of the paternalism that Kant loathes. The other half – and the one that punishment traditionally lacks – is that this state intrusion is motivated by a concern for the citizen's wellbeing. Rehabilitative punishment ostensibly violates this second condition. If rehabilitative interests merely structure the methods of punishing and not its motivation or justification, however, then this would not be the case. Kant could not accept rehabilitation as a justification, but that is no obstacle to his accepting it as the means for selecting the appropriate method of punishment.

This is not to say that there are no obstacles. That punishment can deter enjoys the status of an accepted fact. Whether punishment can also rehabilitate is decidedly less accepted. Although efforts to reform and rehabilitate criminals have been characteristic of the modern penal reforms beginning in the nineteenth century, some might question the possibility of this project. Kant expresses this view in his lectures. Engaging in rational psychology, he claims that "[Punishments] invariably damage morality; the victim believes that if the law had not been there, he would not suffer the physical evil;

thus the law brings about an aversion towards it on his part, and he is thereby hampered from passing free judgments on the morality of his action" (27:556).

Is he correct? The psychological phenomenon he describes is not inconceivable; people might respond in the way he claims. They might not, however, and it is not at all obvious that an individual punished by the state will necessarily, in all cases, experience a degree of resentment that deforms the individual's moral character and obscures her will. It is possible that the correct application of punitive force could serve to aid one in making better choices going forward. This is an empirical question, to be sure, and as such it seems prudent not to rule out either possibility without sufficient evidence.

Even within Kant's lectures, this claim – that punishment can serve to rehabilitate – finds some support. Here, Kant is recorded as saying that "Rewards and punishments can indeed serve indirectly as a means in the matter of moral training," (27:288). He even indirectly provides a picture of how this would work. When describing the role that external factors can play in affecting the functioning of our will, Kant states,

> In general, if the countermeasures are adequate to weaken the inclination, and enliven his sensory feeling by another contrary feeling in collision with it, we are then in a position to ensure that continuing habituation will weaken the power of inclination, and thereafter moral grounds have an impact, so that by removal of the obstacle he is thus made free, and can be brought, by this pathological expedient, to a recognition of his duty. (27:522)

The pathological expedient that Kant references can be any external incentive – includinghe threat or act of punishment. If it is balanced properly, then it can serve as a countermeasure to the inclinations that serve to distract us from what the moral law requires of us. While too strong a sanction might have the effect that Kant discusses

above – namely, causing one to resent the law – a properly balanced punishment could effectively rehabilitate a wrongdoer. Such a punishment would not interfere with the subject's will, but rather remove obstacles that threaten to impair proper use of the will.

There is a further reason to think fundamental elements of Kant's practical philosophy leave room for the possibility of rehabilitation. Kant is famously committed to the view that any person, regardless of past experiences, external circumstances, or apparent character, has the capacity to make moral or immoral choices (5:28-30; 6:50). We all have the capacity for acting morally – a claim that Kant holds to be self-evident – and any record of bad behavior can be potentially reversed. Given this radical freedom, we should conclude that all persons are capable of being rehabilitated. One method by which this could be accomplished is the properly balanced pathological incentive, discussed above.

All of this demonstrates that Kant could allow for the use of rehabilitative methods of punishing, but I still need to show why this would be desirable. The answer to this question lies in the best alternative to *ius talionis*. Rather than fixing the methods or amount of punishment by reference to the law of retribution, Kantian protective deterrence advocates making these specifications in accordance with a proactive understanding of the formula of humanity. Rather than a simple duty of non-interference, we should read the formula of humanity as offering positive prescriptions – albeit, indeterminate ones. Respect for the rational, free personhood of others requires us to not only avoid using them as mere means, but also to attempt to foster in them the conditions necessary for their personhood. What this means will vary from case to case;

most typically, it will require us to help others pursue their ends, provided that these are not inconsistent with morality of the universal principle of right. This obligation must always stop short of the kind of paternalistic interference that actually diminishes the freedom of those we seek to respect. Even if a person believes he knows what is best for his friend, he is not free to make choices for his friend.[8] As a practical point, both individual citizens and the state are rarely in a position to know whether an act of intercession would represent the violation of an individual's freedom or the removal of a pathological impediment to her agency. In punishment, however, we find a unique situation: by engaging in a criminal action, a citizen demonstrates an inability to act in a fully rational and autonomous way.

Ordinarily, the state would not be justified in using coercive force to rehabilitate its citizens' moral character. To do so would be to overstep its authority and contradict its purpose. In the case of punishment, however, these concerns are lessened or removed. According to Kantian protective deterrence, the criminal's actions have already authorized the government to use coercive force against her to ensure the continued deterrent efficacy of the law. While this application of force must still be consistent with the individual's status as an end in herself, the concerns about the state overstepping its purview by evaluating elements of a person's moral character are answered. Further, in the same way that the rational consent of the criminal's ideal, legislative nature can justify the coercive use of punitive force for deterrent ends, it does

[8] As I discussed in chapter four, educating children is a special kind of case in which parents or guardians are not only permitted to make choices for children based on what they believe to be in the best interest of the child, they are obligated to do so.

not seem difficult to imagine that a person could rationally consent to undergoing rehabilitative training if he or she breaks the law.

By adopting rehabilitative methods of punishment,[9] the state could select those practices that best express equal respect for the dignity of all persons, including both the criminal and the victims of the crime. It is important to note, however, that this rehabilitation does not aim simply at producing a specific outcome. If this were the case, rehabilitation could take the form of brainwashing. To 'rehabilitate' in this manner, though, would not demonstrate respect for the personhood of the criminal. Proper Kantian rehabilitation would aim to employ methods that engage the rational aspect of the criminal and provide training that would serve to counteract the pathological impediments to one's acting rightly. Although this would likely result in a person being less likely to commit future criminal acts, it would be for the right reasons.[10]

In his paper "Treating Criminals as Ends in Themselves,"[11] Hill has offers a rough, preliminary sketch of what kinds of changes to the criminal justice system would be necessitated by a more robust incorporation of the formula of humanity. Using this sketch as a starting point, I will offer a partial account of the sort of punitive strategies that express the appropriate respect for the equal dignity of all persons. Hill's paper has a number of objectives, but the section I am most interested in sets aside Kant's

[9] At present, I am only defending the idea of using rehabilitative methods of punishing. I will discuss the obstacles to using rehabilitation to fix the amount of punishment below.
[10] Merle also defends the idea that rehabilitation could be the appropriate way of expressing respect for convicted criminals. See Merle, Jean-Christophe. *German Idealism and the Concept of Punishment*. Cambridge: Cambridge University Press, 2009.
[11] Ibid., pp. 17-36.

professed view of punishment, imagining instead what kind of theory we might expect if we knew only the basics of his moral and political thought. To this end, he writes,

> How much punishment is appropriate to various offenses? From the Kantian legislative perspective sketched above, this would be a very complex question because there are many relevant factors to consider. However, certain over-simple answers are clearly ruled out. For example, we cannot seek answers by a consequentialist cost/benefit analysis that treats all values as commensurable. Nor can we suppose that offenders have a judicially discernible "inner desert" that can be rated on a scale proportionate to the severity of various punishments. The relative effectiveness of deterrent threats would be relevant, but it cannot be decisive by itself because this could authorize punishments that are too severe, or too light, from the perspective that reflects the equal dignity of all persons.[12]

Although Hill's view is only the beginning of a full alternative, it does provide us with an intriguing possibility. In order to guide his deliberations about the appropriate methods of punishment, Hill focuses his use of the formula of humanity on the expression of respect. He considers the various pitfalls that practical policies, such as mandatory fixed sentences for crimes,[13] would face in light of the way in which they succeed or fail to express respect for the "equal dignity of all persons."

One might ask why this particular case is one in which he should strive to actively express respect. After all, the individual toward whom such an expression would be directed is a criminal. If one knew nothing of Kant's established view of punishment, however, this might not seem so strange. When we punish, we inflict some harm in the interest of preserving the deterrent efficacy of the law. The fact that the person who will bear this harm has committed a crime is certainly not an irrelevant factor; a large portion of this chapter sought to establish that this fact plays an important

[12] Ibid., p 30.
[13] Ibid., p 31.

role. Nevertheless, the fact of this person's criminality does nothing to diminish their moral worth as a rational being that is capable of setting ends. Given the situation, it does not seem unreasonable to suggest that an active expression of respect could go a long way toward decreasing the dehumanization of criminals, as well as their animosity toward the law.

This would also be in keeping with at least one understanding of Kant's claim that punishment is itself a categorical imperative. According to Scheid, we should understand the claim that punishment is a categorical imperative as applying to fixed rules of distribution, rather than a strict need for retribution. Scheid offers the following explanation for Kant's confusing and controversial claim:

> Some people have taken this to mean that punishment, as such, is imperative, apart from its consequences. Just before this passage, however, Kant claims the criminal's innate personality protects him against being manipulated merely as a means; and Kant urges against reducing the punishment in a particular case so as to gain some utilitarian advantage. What Kant implies is that the law governing legal punishment is a categorical imperative against using utilitarian considerations to adjust punishments in particular cases. It is clear from the full passage that Kant is talking about the allotment of individual punishments. Again, the point is that judges must not take utilitarian considerations into account when deciding sentences. Hence, the "categorical imperative" refers to questions of distribution or allotment and is addressed to the judges within the system of legal punishment.[14]

It is precisely this concern for expressing respect the equal dignity of all persons, I posit, that can serve as the rationale for using rehabilitative means of punishing. If we accept that persons have the capacity for moral improvement and that punishment, if properly balanced, can accomplish this improvement, then true respect for these individuals as ends in themselves should require us to facilitate this improvement. Note

[14] Scheid, pp. 279-280

that this would not represent an interference with the free will of the criminal; it would merely be the balancing of one pathological incentive with another.

This display of respect for the dignity of the criminal's moral nature by removing pathological impediments would not, however, extend to other citizens; the state's acting in such a way would be far too great of a violation of individual freedom. For individuals who have committed crime, the state's interests already require using them to further the deterrent efficacy of the threats of legal sanction – a use to which the citizens rationally consent, as colegislators. Given that they will be used in this way, the state has already begun making decisions on their behalf. It seems to be a better expression of respect for the dignity of the criminal to deter future crimes through rehabilitative efforts than to merely cause suffering as a way of frightening others.

These rehabilitative methods can also better express respect for the dignity of those harmed or wronged by the crime committed than retributivism. For the retributivist, the way of expressing respect for these persons is causing the perpetrator to suffer. With rehabilitative methods, however, punishment becomes a tool for eliminating the circumstances, motivations, and maxims that made the crime possible in the first place. While some victims might feel better respected by the state's retributive harming of the guilty, this comes across more as an exclusive desire for vengeance.

By instituting an equal regard for the rational humanity of each criminal being punished, we could guarantee not only proportionality, but address some of the concerns about equality and stability that troubled the deterrent approach. Although their nature as biological humans might entail differences in their responsiveness to

threats – thus necessitating penalties of varying strengths – attention to these differences would be ignored. Focusing on such features might appear to offer a higher level of respect for each individual, but in reality the differences to which we would be attending are of no greater moral significance than mere inclination. The best way to show respect would be to institute a policy of equal punishments for the same crime, regardless of personal details.[15]

Throughout this section, I have been focusing on the appropriate means for selecting the method of punishing that will be employed. There still remains a question, though, of what amount of punishment is appropriate. Given that I have argued in favor of rehabilitative methods as the best way to express the kind of respect for persons as ends in themselves, it might seem natural to fix the amount of punishment rehabilitatively as well. According to this approach, the correct amount of punishment would be however much it takes to rehabilitate an individual.

While this has some appealing features, I think that this approach to determining and specifying the appropriate amount of punishment faces too many difficulties. Kant is heavily committed to a high degree of agnosticism or skepticism about our ability to understand the motivations and maxims of ourselves and others. Verifying whether or not an individual had been successfully rehabilitated would require an impossible amount of insight into an individual's character and inner psychology.

---

[15] Obviously, certain personal details might still be considered exculpatory. For instance, insanity might still be a reason for decreasing a sentence. This case, however, represents a special class of exceptions. Allowing for such exceptions need not be incompatible with prohibiting the practice of determining penalties based on one's individual receptiveness to the deterrent force of threats.

Instead of focusing on this rehabilitation of character, I posit that we should fall back to deterrence as the means for selecting the appropriate amount of punishment. Criminals would experience punishment designed to rehabilitate, but rather than specifying the appropriate amount as whatever produces a change in the criminals' character, the amount would be that which deters further crime. These two amounts might ultimately be the same, but they need not be. We might worry that it is just as difficult to determine whether an individual has been successfully deterred as it is to determine if he has been successfully rehabilitated. I do not think this is so, for two reasons. The first is that deterrence has both individual and vicarious functions. Even if we have difficulties with the first, success at the second could still be an effective way of determining the appropriate amount of punishment. Second, it is easier to observe regularities in the amount of penalty that will be required to deter crime than it would be to try to create general policies about how much punishment it would take to rehabilitate. As long as some target percentage of the population was deterred, then we would know we had the correct amount of punishment.

This cooperation between rehabilitation and deterrence also has the virtue of capturing the spirit of Kant's practical philosophy. The rehabilitative methods of punishing serve well to fulfill the justifying purpose of punishment, namely the deterrence of future criminal actions. It does so in a way that comports with the duty to express respect for the equal dignity of all persons that stems from the formula of humanity. Yet, at the same time, it refrains from overly paternalistic concern with the character of the citizens. So long as they are effectively deterred, the state's interest in

punishing is satisfied; no further punishment is to reform their characters. As Kant writes, these policies should be sufficient to maintain a rightful condition, even amongst a nation of rational devils (8:366).

6.5 Execution

Perhaps the most infamous aspect of Kant's theory of punishment is his endorsement of the permissibility – indeed, the necessity – of capital punishment. Given the frequency with which Kant uses murder as his primary example of crime, one might mistakenly come to the conclusion that Kant favored executing criminals for all manner of offenses. While this is not the case, there are crimes other than murder that he believes merit execution, including – but not limited to – treason and the attempted or successful assassination of political officials (6:320). Although he does not suggest a method by which the condemned ought to be executed, it seems clear from his stance on torture that it ought to be swift and cause no more pain or suffering than is necessary. Death, rather than misery, is the sentence and the goal.

Kant's support for capital punishment is based upon moral desert, as specified by *ius talionis*. Given the difficulties that led me to reject *ius talionis* as a suitable principle by which to fix the methods and amount of punishment, there is no reason why Kantian protective deterrence must hold execution to be strictly necessitated in the kinds of cases that Kant describes. Just because it is not required, however, does not meant that execution is always impermissible. In order to make some stronger claim about the

absolute prohibition of capital punishment, I will need to go beyond simply showing that Kant's reasons for supporting it do not apply to Kantian protective deterrence.

In this section, I will endeavor to do precisely that. I will argue that execution is inconsistent with Kantian protective deterrence, on the grounds that executing fails, in almost every case, to respect the humanity of the target. Building off of the same arguments I employed above to defend rehabilitation, I contend that genuine respect for the autonomy and humanity of a person is incompatible with ending his or her life. Although I will consider a small class of cases in which it is possible to kill a person while still holding him or her as an end, I will argue that these cases are not of the right sort to ground even a limited legal use of capital punishment.

Given the deterrent justification of punishment I have been defending, the most obvious place to object to the use of capital punishment might be to cast doubt on its effectiveness at deterring future offenses. While capital punishment is inarguably effective as a deterrent against future crimes being committed by the one being executed, there are other ways to guarantee the same outcome without resorting to execution (such as lifelong imprisonment). Furthermore, capital punishment's effectiveness as a vicarious deterrent is highly suspect, with numerous empirical studies demonstrating little to no effect on the commission of crime.[16] Even if execution does

[16] See Radelet, Michael L. and Lacock, Traci L. "Do Executions Lower Homicide Rates?: The Views of Leading Criminologists." *The Journal of Criminal Law and Criminology*. Vol. 99, No. 2 (2009), pp. 48-508. There are opposing studies that do purport to show a deterrent effect; See, Dezhbakhsh, Hashem et al. "Does Capital Punishment Have a Deterrent Effect? New Evidence from Postmoratorium Panel Data." *American Law and Economics Review.* Vol. 5, No. 2 (2003), pp. 344-376; and Ehrlich, Isaac. "Capital Punishment and Deterrence: Some Further Thoughts and Additional Evidence." *Journal of Political Economy*. Vol. 85, No. 4 (August 1977), pp. 741-88.

have some deterrent force, we might conclude that it is insufficiently effective to justify the loss of life. For all these reasons, a deterrent theorist might argue that empirical contingencies would prevent execution from being justified as an appropriate form of punishment.

Although these are all legitimate concerns, they are most likely insufficient to demonstrate the absolute impermissibility of execution as a punishment within the context of Kantian protective deterrence. Recall that I have been advancing a mixed theory. While deterrence serves as the justification, the other elements of the theory are not merely intended to maximize the deterrent force of legal prohibitions. Execution could be ruled out if it were directly counter-productive to the deterrent justification, but it is unlikely that this strong a claim could be empirically demonstrated and defended. Thus, even if capital punishment is not the most deterrent possible option, it could still be permitted.

Simple deterrence, however, is not the only possible reason to reject execution. Instead, we should focus on the act of killing itself and the mindset of the executioner. There is a substantial literature addressing the question of Kant's commitment to capital punishment, representing a number of different points of view.[17] Although there is no consensus, there is a common thread running through some of the scholarship: the act of

[17] Altman, Matthew C. "Subjecting Ourselves to Capital Punishment: A Rejoinder to Kantian Retributivism." *Public Affairs Quarterly*, Vol. 19, No. 4 (Oct., 2005), pp. 247-264; Ataner, Atilla. "Kant on Capital Punishment and Suicide." Kant Studien. Vol. 97, No. 4, pp. 452–482; Yost, Benjamin S. "Kant's Justification of the Death Penalty Reconsidered." *Kantian Review*, Vol. 15, No. 2, 2010.

killing – or a law requiring the same – seems to be impossible to morally carry out against anyone that the killer truly regards as a person.[18]

The basic points of this view are grounded in Kant's moral philosophy. As a bearer of rationality and autonomy, a person should never be treated as a thing (4:228). Part of what this means is that the wills of others should have a direct or indirect effect on the way we behave; we ought not to behave in ways that could not be universalized, and we should not treat others in ways to which they could not morally consent. Furthermore, a person could never morally consent to her own death. Thus, the act of killing another must always be a moral wrong; to engage in it is to treat the victim as a thing, rather than a person.[19]

One might be tempted to say that this argument would rule out the possibility of punishing in any way. After all, every time someone is punished, it is mostly likely

---

[18] Self-defense seems like a necessary exception to this claim. In instances of self-defense, Kant holds that you could conceivable kill another person without acting in a way that fails to respect him as an end in himself. Doing so is only permissible, though, because it actively hinders a hindrance to freedom – namely, one's own. In the case of execution, it is no longer the case that death is an active hindering to a hindrance to freedom. While the threat of capital punishment might be construed in this way, I find this too indirect an explanation. A person could consent to a law whose sanction threatens to use any violator – including herself – to deter others, but she could not consent to die for this purpose.

[19] There is also reason to think that Kant should object to execution on grounds of concern for its effect on the *executioner*. When considering whether we might have obligations to animals, Kant concludes that despite the impossibility of a direct duty not to harm non-rational animals, we nevertheless ought to refrain from needless violence, due to what such behavior does to us (27:710). He argues that wanton cruelty to animals can warp our moral sensibilities, making us more prone to violence. Evidence suggests that those who participate in executing convicted criminals are more prone to mental health issues, including violent behavior and suicidal tendencies. While engaging in activities that lead to such outcomes might violate a duty to oneself, I think there are ultimately much stronger and more obvious ways to demonstrate the impermissibility of suicide within Kant's practical philosophy. For details about secondary trauma, see Gil, Amanda et al. "Secondary Trauma Associated with State Executions: Testimony Regarding Execution Procedures." *Journal of Psychiatry and Law*. Vol. 34 (2006), pp. 25-36.

against his will. This is not the case, however; as I have argued above, criminals can and do will their punishment in a rational, hypothetical sense. Death, however, cannot be willed in this same manner. To understand why death is particularly something to which a person could never consent, we should look to what Kant writes in the *Groundwork* about impermissibility of suicide:

> If he destroys himself in order to escape from a trying condition he makes use of a person *merely as a means* to maintain a tolerable condition up to the end of life. A human being, however, is not a thing and hence not something that can be used *merely* as a means, but must in all his actions always be regarded as an end in itself. I cannot, therefore, dispose of a human being in my own person by maiming, damaging, or killing him. (4:429)

Kant's reasoning is that any person who commits suicide destroys himself in order to achieve some end, and in so doing, uses himself as a means to that end. There's good logical and textual reason to think that this is not always the case. Indeed, ending one's own life can – in certain cases – be an expression of respect for autonomy. In Collins's lecture notes, Kant discusses the case of Cato the Younger, who famously committed suicide in order to preserve his dignity and help protect republican Rome. Kant speaks positively of this decision, highlighting the fact that Cato was to be put to death by Caesar and that his choice to end his own life was taken only because of the great effect that his suicide would have in preserving the freedom of his nation (or, at least, so Cato hoped). While suicide to preserve the freedom of a nation is presumably quite rare, the fact that taking one's own life might be permissible in certain cases indicates that it is conceptually possible to hold a person as an end even in the act of killing him or her.

This possibility is connected to the reason that Kant takes life to be valuable. He states that "in and for itself, life is in no way to be highly prized" (27:372); preserving our lives, then, is foremost amongst our duties to ourselves for reasons other than our mere biological existence. Elsewhere in Collins's notes, Kant gives us an indication that what we must preserve is our freedom (27:342-334). Suicide is wrong because it destroys all future freedom. Only by living, even through great hardship, can we show an adequate degree of respect for the rational autonomy of our person and act in accordance with the greatest freedom for all. This also comports with the example of Cato: in his case, his autonomy was greatly constrained, and he recognized that it would be extinguished altogether in the near future. By committing suicide, he did not eliminate any future freedom; instead, he made the last free choice that was available to him in the hopes that it would inspire greater freedom and autonomy.

To sum up, suicide is typically prohibited, as it involves willing the destruction of one's own autonomy and humanity. While properly recognizing the value of these features of ourselves, it is impossible to also act in such a way as to end one's own life. By willing death, one wills an abrogation of one's freedom, which is both inconsistent with other universal goals of human beings and lacking in appropriate respect for the capacity that makes humans into moral persons. Only in cases in which this freedom is already lost can a person permissibly will her or his own death. For instance, a person facing a drastic loss of autonomy might permissibly end her own life, if doing so was a final act of respect for her own humanity. Even if the fate of the republic does not hang

in the balance, the certain prospect of debilitating disease could serve as a sufficient reason to use the final exercise of one's autonomy to end one's own life.[20]

If this is right, then it has implications for the permissibility of execution. Executing another – like killing oneself – is typically inconsistent with the moral imperative to treat that person as an end in herself. The example of Cato or the person who kills herself in the face of debilitating terminal illness shows that it is sometimes possible to end the life of an autonomous being while still holding that being as a person worthy of respect. While this might carve out room for permissible instances of euthanasia, it is unlikely that it could ever be sufficient to ground capital punishment. . would be practically viable. There are numerous reasons why it seems wither

We can conclude, then, that capital punishment could not be established as a legal possibility within Kantian protective deterrence. This is another instance in which I am compelled to reject outright some of Kant's own statements. These statements, though, are ones that seem deeply irreconcilable with the very foundations of Kant's moral philosophy. Given his commitment to the view that all people are capable of reform, no matter how badly they have acted in the past, it seems unconscionable for anyone to execute another person. Doing so requires one to reject the status of the victim as deserving of respect.

[20] This would also cover ending someone else's life, under similar circumstances. While mercy killing would still remain impermissible, physician-assisted suicide would presumably become permissible.

Conclusion

There is a certain purity to the strict retributivism that Kant advocated in his writings on practical philosophy. As an advocate of deep retributivism, all the elements of his theory fit together with a kind of singular harmony. Once it becomes clear that Kant does not have the means to explain why the state would be justified in responding to this moral desert, however, the purity of this account becomes its own downfall. *Ius talionis* is compelling in its apparent simplicity, but upon closer examination, any number of fatal flaws undermine its viability as a principle by which a state could settle upon the appropriate methods and amount of punishment. Without Kant's strict retributive justification, there remains no good reason to try to rehabilitate the literal understanding of *ius talionis*.

Instead, it is necessary to seek a less literal way to satisfy Kant's concern for proportionality. Although the traditional way of solving these issues – the fitting of crimes to punishment based on their relative positions on the 'scale' – does not avoid all the problems, I have argued that there is hope. Specifically, by focusing on rehabilitative methods of punishing, proportionality in punishment becomes more focused on respecting the humanity of the criminal by committing to fostering the conditions of her rational agency. This commitment to punishment as a way of respecting the autonomy of the criminal by trying to foster within her the conditions for free and rational choice is a way of further constraining the dangers of run-away deterrence that Kant fears, and it is deeply harmonious with his Formula of Humanity.

7 *Quis Custodiet Ipsos Custodes*?: Resisting and Punishing State Authority

The striking contrast between Kant's personal enthusiasm for the American and French revolutions and his strict, near-authoritarian political philosophy has been extensively documented in recent years.[1] While many of these papers focus on demonstrating the underlying consistency of his simultaneous condemnation of rebellion and praise of its effects, I intend to take a different approach. Given our focus on punishment, it is my aim to examine what role the institution of punishment plays in Kant's prohibition of revolution or punishing former state authorities. While we have explored the legitimate ways in which that authority is meant to function and when it is authorized to coerce, we have yet to consider what remedies exist in Kant's system for authorities that exceed or abuse these limits of legitimacy. This is an especially pertinent issue, as Kant explicitly connects the authority that such figures wield with the use of coercion and the determination of right in cases of legal dispute.

As such, it is worth asking: what of the legislative and executive authorities themselves? In what sense or under what conditions, if any at all, can their power be resisted? Does such resistance always merit punishment? Could there even be cases in which resisting a law is morally required of us? How can these authority figures be held

[1] See Hill, Thomas E. "Questions about Kant's Opposition to Revolution." *The Journal of Value Inquiry*. Vol. 36, No. 2-3 (2002). pp. 283-298; Nicholson, Peter. "Kant on the Duty to Never Resist the Sovereign." *Ethics*. Vol. 86, No. 3 (1976). pp. 214-230; and Reiss, H.S. "Kant and the Right of Rebellion." *Journal of the History of Ideas*, Vol. 17, No. 2 (1956). pp. 179-192.

accountable by the citizens? Likewise, is it permissible for them to be held accountable by the leaders of other states?

This constellation of questions is unified by a central theme: namely, they all investigate the role that authority plays in the civil institution of punishment. Kant's opinion on these matters is similarly unified. In the "Doctrine of Right," he definitively states "There is…no right to sedition, still less to rebellion, and least of all is there a right against the head of a state as an individual person, to attack his person or even his life on the pretext that he has abused his authority" (6:320).[2] The reason for his staunch denial of such rights is his commitment to the necessity of determinate answers in cases of conflict. The law exists, in part, due to the necessity of having some authority to settle matters of dispute between parties. In the state of nature,[3] there is no possible mechanism for placing others under an obligation to respect our use of external objects, and thus we will perpetually come into conflict with others with whom we come into contact. We need the law to solve this problem, and the law only functions when it can give determinate answers in all cases of dispute. As we will see, Kant envisions both the executive and legislative as protected by variations of this argument.

In the foregoing chapters, I have focused on an interpretive reconstruction of Kant's theory of punishment. Throughout, there has been an emphasis on building a Kantian theory of punishment that 1) is consistent with his most foundational

[2] All internal citations can be found in Kant (1996).
[3] NB: Kant does not believe the 'state of nature' to be a historical state of human development. Rather, he envisions it merely as a hypothetical scenario, a useful tool for determining what kinds of institutions people would agree to.

philosophical commitments and 2) preserves as many of his statements about punishment as possible, where this does not violate 1). In this chapter, I apply this same methodology to some of the practical implications for how Kant imagines the institution of punishment will be instantiated in civil society. According to Kant's thinking, the possibility of punishment requires both law and the possibility of enforcement; in other words, there must be a legitimate legislative power to create the laws whose violation creates the possibility of punishment, and there must be a designated executive authority that bears the sole right and responsibility to carry out such punishments. In one sense, we can be guaranteed of this division of government by the necessity of punishment; Kant thinks the state's authority is irresistible for similarly necessary reasons.

I contend that in following Hobbes and the traditional currents of political thought so closely, Kant makes the converse of the mistake he made with respect to punishment. As we determined in chapter four, Kant's support for a retributive theory of punishment fails to offer any successful, justificatory arguments. In that case, his radical aspirations were foiled by the underlying conventional foundations of his political philosophy. The problem with Kant's absolute denial of the permissibility of civil disobedience, rebellion, or punishing previous rulers comes not from conventional underpinnings, but rather from his own moral philosophy; in this case, his conventional aspirations are foiled by the radical force exerted by his moral philosophy. Put plainly, while Kant can successfully argue against a legal right to civil disobedience, resistance, or revolution, his efforts to show a moral obligation to refrain from such rebellious

actions do not and cannot succeed. For similar reasons, his argument against the punishing of former authority figures is similarly unsuccessful.

In making this argument, I will be defending an interpretation of Kant's legal philosophy that could be described as 'constrained positivism.' Like an orthodox positivist, Kant holds that the merits of a law are to be determined by their creation in a fixed legal procedure, rather than by an appeal to some external standard. Unlike a fully positivist legal theory, however, Kant takes there to be several strict limitations on what can become law. Significantly, if a proposed policy fails to satisfy the necessary requirements of law, this does not make it a bad law; instead, the policy *is and can be no law at all*. It is precisely this limitation that will enable a Kantian form of civil disobedience and active resistance to state power. While we are morally obligated to follow laws, we are permitted – and perhaps even required – to disobey, refuse, and resist policies that cannot be legitimately legislated.

7.1 Opposing State Power

Throughout this section, I will be distinguishing three types of legal disobedience and resistance. The first is civil disobedience. I use civil disobedience to refer to a passive refusal to comply with a single law or small cluster of laws. Those engaging in civil disobedience do not resist arrest, and they do no break other laws in protest. According to this usage, staging sit-ins at a lunch counter in order to protest laws restricting access to such restaurants would be civil disobedience; staging a march without a license,

though, is only civil disobedience with respect to the law requiring permits for marches. The goal of civil disobedience is to change a law or set of laws or policies.

The second form of legal violation I will be discussing is resistance. Resistance is distinguished from civil disobedience in two ways. First, involves breaking a wider range of laws. Second, resistance involves a refusal to submit willingly to arrest and punishment. Resistance does not involve launching attacks on the police or other government officials, for it does not aim at the overthrow of the state. Instead, its goal is to bring about substantial changes to some agency or wide-sweeping set of policies.

Finally, Rebellion or revolution occurs when the aim is to replace the government with a new one. Rebellion involves actively attacking the forces of the government; it is essentially a war declared on the state as it currently exists. If a person or group is engaged in rebellious actions, they no longer accept any of the state's laws or authority figures as legitimate.

Before proceeding further, I should say a word about the form of republican separation of powers that Kant supports, as the appropriate response to executive and legislative abuses of power might potentially differ.[4] In some places, Kant provides passages that seem to blur the distinction between the various branches of government (such as a key example at 6:321 that I will be discussing shortly). This could be due to simple error, to an accidental confusion over his own terminology, or to a desire to avoid

[4] Throughout this chapter, I will be addressing only the rights of citizens who live in a state that creates, previously created, or comes very close to creating a rightful condition. As a rightful condition can only be truly acquired and maintained under a republican government, my focus will be on citizens who live in republican states.

again angering Frederick William II, who had already censured Kant's writings on the subject of religion.[5]

Regardless of whether the occasional lack of clarity is due to error or self-preservation, it is clear from 6:313 in the Public Right section of the "Doctrine of Right" that he holds supreme sovereign authority to rest with the legislative branch of government. The legislature represents the united will of the people—the only possible source of political legitimacy. Legislative authority can rest with either a single law-giver or with a legislative body, such as a senate. These claims of legislative sovereignty are more or less in keeping with the post-Hobbesian social contract tradition, as well as the natural law tradition; in particular, the prioritizing of legislative power as an expression of the 'general will' has a distinctively Rousseauian character to it.[6]

This generality is the first of two conditions for legitimacy of a law that Kant outlines in the *Metaphysics of Morals*. Laws that arise from the legislative branch must be general in two senses. First, a law must be general in its content (6:316-6:317). Only policies that refer to the whole people or some broad group of citizens – rather than to particular individuals – and that are intended to serve as fixed, exceptionless rules that

[5] Beck makes a compelling case against viewing any of Kant's inconsistencies as strategically motivated. As he observes, "While it is not improbable that Kant was intimidated by the censor, I find it incredible, for Kant's actual response to the censor in 1792 was silence, not deception. In 1766, he had written Moses Mendelssohn, "Although I am absolutely convinced of many things that I shall never have the courage to say, I shall never say anything I do not believe." I think that was as true in the 1790's as in the 1760's; and therefore, I must try to find some other way to explain the apparent inconsistency in Kant's attitudes." (See Beck, Lewis White. "Kant and the Right of Revolution." *Journal of the History of Ideas*. Vol. 32, No. 3 (1971), p. 411)

[6] See 6:314: "Therefore only the concurring and united will of all, insofar as each decides the same thing for all and all for each, and so only the general united will of the people, can be legislative."

do not conflict with other such rules can become laws. Second, laws must be general in the sense that they "involve the unity of and resolution of conflicts in accordance with universal laws."[7] Only when a policy can satisfy the requirements of universal law can it become a civil law, and policies that spring from the particular wills of individuals cannot be guaranteed to reach this standard. It is this second kind of generality that grounds the first; the only kind of law that can result from the subsuming of individual wills under the dictates of universal law are those that have a general content.

The second condition for legitimacy is rational or hypothetical consent. Only those policies to which all citizens could possibly give their rational consent can be made into law. It is possible for a law to still be legitimate if one or more citizens do not, in fact, give their consent, if their refusal is based on some irrational inclination. If even one citizen has a rational basis for rejecting a law, however, then this is sufficient to render the law illegitimate, and thus nullify it as a possible law. In "On the Common Saying", he writes "What a people cannot decree for itself, a legislator cannot decree for a people" (8:304). The legislative does not merely act wrongly if it attempts to institute such a policy; it does that which it does not have the power to do. For an example of something that a person cannot will, we can look to the Doctrine of Right at 6:329-6:330 where Kant describes how a person cannot possibly will herself or himself into slavery. He writes, "Since we cannot admit that any human being would throw away his freedom, it is impossible the general will of the people to assent to such a groundless prerogative, and

[7] Mulholland, p. 301.

therefore for the sovereign to validate it" (6:329). Thus, any law that relegates a citizen to a position of servitude would fail the second test and thus be illegitimate and beyond the power of the government to legislate or enforce.

The executive branch, on the other hand, is responsible for the implementation and enforcement of laws, the execution of punishments for any violations of the laws, and all other institutions involved in the day-to-day operations of the state (e.g., the recording of contracts, deeds, etc.). The executive head of state, to whom Kant refers as the 'ruler,' is the agent of the legislative; he or she has no authority except that which is derived from the power the legislative bestows upon him or her (6:316). The policies of the executive are 'decrees,' not laws, and as such they can and must be particular. It is important to note, though, that the executive has wide latitude in determining the parameters of how laws will be enforced; while the letter of the law and even the specific punishment warranted by its violation are spelled out by the legislature, all decisions about how to enforce the laws are determined by the executive; the legislature lacks the ability to directly check individual measures of the executive. As such, the executive could enforce a perfectly legitimate law in a way that violates the rights of the citizens.[8] The only power the legislative has to curtail the decrees of the executive is to pass a new law or replace the executive with a new agent.

[8] An example of such a scenario might be a law giving the police the power to search vehicles pulled over for routine traffic violations for illegal narcotics. While such a law might be legitimate, we could imagine a scenario in which the executive elects to only exercise such a power when dealing with certain racial minority groups. In such a scenario, we might think of the executive as enforcing a legitimate law in an illegitimate way.

Taken at face value, Kant's rejection of any right to resist that authority of the state does not seem to be in any way affected by this distinction between the sovereign, legislative power and the subsidiary, executive power. Although he recognizes the difference between these two branches, he holds that resistance to either one is strictly impermissible. Of the legislative's imperturbable supremacy, he writes:

> The reason a people has a duty to put up with even what is held to be an unbearable abuse of supreme authority is that its resistance to the highest legislation can never be regarded as other than contrary to law, and indeed as abolishing the entire legal constitution. For a people to be authorized to resist, there would have to be a public law permitting it to resist, that is, the highest legislation would have to contain a provision that it is not the highest and that makes the people, as subject, by one and the same judgment sovereign over him to whom it is subject. This is self-contradictory, and the contradiction is evident as soon as one asks who is to be the judge in this dispute between people and sovereign. For it is then apparent that the people wants to be the judge in its own suit. (6:320)

This is the core of his objection to resisting state authority. In Kant's political philosophy, all rights are claims that citizens have against other citizens. These claims are guaranteed by the authority of the state. Put another way, if I violate another citizen's right, the citizen is entitled to the state's use of coercive force to recoup whatever losses were sustained as a result of my action. Given their connection to state enforcement, rights cannot exist outside of a rightful condition (6:311). While certain moral duties exist independently of a rightful or juridical state, these are exclusively the unenforceable ethical duties that we all have as free, rational persons. Rights, on the other hand, can only exist in a civil society that enjoys both the rule of law and a determinate power who has the authority to enforce the law.

In order for citizens to have any legal right to actively resist the implementation of a law, there must be another law that extends this freedom to them and guarantees their exercise of it; they must have a legal claim that can be enforced by the state's coercive power. Any law that extends to a people the right to disobey the law whenever they see fit is both highly impractical and, more importantly, contradictory. How could we make sense of a legal right that would require the state to defend, with force if necessary, a citizen's entitlement to resist the power of the state? If this were the case, each citizen would have the power to command the state to alter or fail to enforce any law at any time.

The legal contradiction that Kant sees as prohibiting any constitutionally recognized right to insurrection is also grounded on a moral contradiction. Recall that all legitimate laws must be passed by legislative action that occurs in accordance with the general, collective will. In light of this, Kant claims that all laws that are passed by the legislative are ones that each citizen has individually willed. The law that requires me to respect my neighbor's property is not an alien constraint, but rather one that originates within my own will. For me to break such a law clearly involves a contradiction, but to Kant's mind, so too does my resisting any law. In resisting, I claim that I simultaneously will a law and do not will the law. Howard Williams writes,

> From a moral point of view the State represents the general will of the people, and the individual citizen must see himself as part of this general will which creates the law and brings into being the sovereign who it is his duty to obey. For the individual to rebel against the State is, therefore, from the moral viewpoint, for him to rebel against himself, and this, Kant argues, is impossible.[9]

[9] Williams, Howard. *Kant's Political Philosophy*. New York: St. Martin's Press, 1983, p. 200.

Of course, this presumes that I have or could have, in fact, willed the law in question. I might privately disagree with what I could rationally will, but I do not have legal standing to dispute this, as there is no one to adjudicate this dispute. Thus, my only recourse as a citizen is to express my opinion through the legitimate, legal channels and, in the meantime, accept whatever answer the legislative authority settles upon.

Although he rules out any active resistance against the state, Kant does seem, at times, to allow for the citizens to passively refuse to comply with a law that would require them to engage in immoral behavior. He has been read this way by numerous interpreters,[10] and there is some evidence for such a reading; after all, Kant does describe a people that always complies with any command from the executive as "corrupt" (6:322). Such readings, however, overlook that in both "On the Common Saying" and the *Metaphysics of Morals*, Kant is specifically referring to a right that the legislators retain. The legislators are the ones who are meant to refuse immoral commands of the executive, and any right to resist that the people have is conducted through their legislative proxies. In other words, Kant specifies that such passive resistance is afforded only to the citizens who are members of parliament (8:297, 6:322). If this is the correct reading, then it is the legislative branch that can passive resist the power of the executive;[11] the people, in this case, have no legal right to resist the state's authority. For individuals to do so would be for "each resistance [to] take place in conformity with a

[10] See Williams and Reiss, H. S. "Kant and the Right of Rebellion." *Journal of the History of Ideas*. Vol. 17, No. 2, Apr., 1956.
[11] For support for this view, see Guyer, p. 289.

maxim that, made universal, would annihilate an civil constitution and eradicate the condition in which alone people can be in possession of rights generally" (8:299).

I think this is the correct way of understanding Kant's position. Extending to the citizens a legal right to passively refuse to obey a law would result in the same problems Kant sees in recognizing a right to actively disobey or resist the law. As such, we ought to read Kant as prohibiting even a guaranteed right to civil disobedience.[12] Pointing out this issue, Mulholland writes, "A right to do as conscience dictates would allow everyone to do as conscience dictates on all matters, including questions of conflict over rights, and even when the objective judgment is mistaken. Indeed, it would allow coercion of the state whenever conscience dictated that this would be the right thing to do. But such a right would make a civil condition impossible."[13] He goes on to assert that despite this lack of legal right, citizens should be morally entitled to a passive refusal to obey a law. I will return to this point a little later, arguing that the moral permissibility that Mulholland recognizes should extend considerably further than mere passive refusal.

The executive power, although merely an agent of the sovereign, is no less unassailable in its authority. Kant writes,

---

[12] One might make an interesting case for the permissibility of civil disobedience in the same manner as Rawls does in his paper "The Justification for Civil Disobedience." Rawls famously defends civil disobedience as an act of political speech, intended to address some injustice and bring about a change in policies or institutions (see Rawls, John. "The Justification of Civil Disobedience." *Collected Papers*. Cambridge: Harvard University Press, 1999, p. 181). Given Kant's strong commitment to the importance of freedom of speech in a juridical state (see, for instance, 8:304), this might be an approach that could gain some traction with Kant's underlying political philosophy.

[13] Mulholland, p. 339.

> The sovereign has only rights against his subjects and no duties (that he can be coerced to fulfill). – Moreover, even if the organ of the sovereign, the *ruler*, proceeds contrary to law, for example, if he goes against the law of equality in assigning the burdens of the state in matters of taxation, recruiting, and so forth, subjects may indeed oppose this injustice by *complaints* but not by resistance. (6:319)

The executive, in other words, is also immune from opposition. Although the citizens have the right to work within the system to bring about changes in the executive's leadership or policies, they cannot go beyond the established channels of registering their discontent. Despite the similarities, though, the reason for the executive's irresistible power is slightly different from the reason why the legislative authority cannot be opposed. Instead of focusing on the contradiction that arises from allowing private wills to oppose the laws that are produced by the general will, Kant's defense of executive irresistibility highlights the contradiction that arises from challenging the structure of legal right and coercion. He writes,

> Even the constitution cannot contain any article that would make it possible for there to be some authority in a state to resist the supreme commander in case he should violate the law of the constitution, and so to limit him. For, someone who is to limit the authority in a state must have even more power than he whom he limits, or at least as much power as he has; and as a legitimate commander who directs the subjects to resist, he must also be able to *protect* them and to render a judgment having rightful force in any case that comes up; consequently he has to be able to command resistance publicly. In that case, however, the supreme commander in a state is not the supreme commander; instead, it is the one who can resist him, and this is self-contradictory. (6:319)

As we can see, Kant thinks the executive must be obeyed because of its connection to external right. In order for there to be a sovereign, there must be a single, determinate individual or office that holds the power to execute the law, through the use

of force, if necessary. If the citizens are capable of preventing the execution of particular laws, then each becomes sovereign in a very real sense. As Williams observes,

> A state which possessed a constitution which allowed the citizen always to criticize and overturn the acts of a sovereign would be thoroughly ungovernable. Depending on the way one wished to look upon it, it could either be said to possess two sovereigns or no one at all. Under such a constitution, both the ruler and the subject would be sovereign. This kind of constitution Kant describes as nonsense.[14]

While resisting the legislative branch would be disastrous in that it would eliminate the possibility of law, resisting the executive branch would also lead to the dissolution of the juridical state by making the enforcement of laws impossible. As both law and someone with the power to enforce it are necessary conditions of a rightful state, resistance of this sort would make a republican state unworkable.

We should not conclude, though, that the above quotations imply that the executive's abuses must be tolerated by the legislative as well. While the people, as subjects, must respect and obey the executive's authority, the legislative sovereign still has the power to revoke the executive's power, remove her from office, and replace her with a new agent. As we will see in the next section, Kant holds that even in the event that such a replacement of the executive is necessary, this does not entitle the state to punish the former ruler.

In the event that the executive refuses such an order, on the other hand, then he loses the authority to act as the state's ruler. Instead, the former executive would become an enemy of the state. The citizens would be entitled to resist the actions of such a rogue

[14] Williams, p. 201.

figure based on their right to self-defense. In all likelihood, the legislative would appoint a new executive figure, whose first order of business would be to subdue her predecessor. In such a case, it is possible that the citizens would be enlisted in the effort to pacify the former executive, but their actions would not be constrained as resistance or rebellion, as they would be acting in accordance with the decrees of the new head of state.

The case of the rogue executive gives us insight into the only possible case of morally acceptable resistance that Kant considers. Much has been made of Kant's historical support (at least, initially) for the American and French revolutions, offered in correspondence and his earlier work. In the *Metaphysics of Morals*, Kant suggests that the initial actions of the French rebels during the revolution of 1789 were potentially justified, not by legality or even morality, but by necessity. The state had devolved to such a condition that it no longer represented an actual civil society; the citizens had, at some point, ceased to be members of a people and had instead found themselves plunged back into the state of nature. In any situation where the legislative can no longer make the claim that it is representing the united will of a body of people, it no longer has rightful authority over them. Note that this situation does not give the former citizens legal or even moral title to oppose or overthrow those exercising coercive power over them (the former authorities), but merely a right of necessity. This is presumably the same kind of right of necessity at work when a survivor of a shipwreck forces another survivor off of the plank of wood that can only support a single individual. Unfortunately, Kant gives us no clear guidelines for determining at what point the state

is so chaotic and divided so as to revert to a state of nature. We cannot be entirely sure what failures or tyrannical behavior on the part of the legislative it would take to remove their legitimate legal authority and open this right of necessity to revolt. Also, given that one can act in accordance with a right of necessity and yet still be described as acting impermissibly, it seems prudent to keep our focus exclusively on the legal and moral arguments Kant offers against revolution.

We have reached a largely complete picture of Kant's position on resistance and rebellion. There can be no legal right of any kind to civil disobedience, resistance, or rebellion. As Thomas Hill sums it up, "Kant argues that trying to incorporate an alleged right to revolution into a constitution for a legal system would be incoherent because it would purport to be a legal authority to destroy the very source of legal authority. Someone cannot coherently claim legal authorization to overthrow the highest legal authority. This seems undeniable."[15] Furthermore, such actions are morally impermissible. Engaging in them violates our moral duty to obey the law (by contradicting our rational willing of the law) and threatens the existence of the juridical state to which we have an obligation to belong.

It is worth noting that this description of Kant's rejection of resistance or revolution is largely drawn from the *Metaphysics of Morals*. In "On the Common Saying," Kant offers a slightly different argument against revolution, which Kant's interpreters have almost universally found unsatisfying. Rather than the formal arguments about

[15] Hill, p. 189.

contradictions that Kant uses later, in "On the Common Saying" he defends his position by reference to the impermissibility of revolution motivated by a concern for happiness. He writes,

> Thus if a people now subject to a certain actual legislation were to judge that in all probability this is detrimental to its happiness, what is to be done about it? Should the people not resist it? The answer can only be that, on the part of the people, there is nothing to be done about it but to obey. For what is under discussion here is not the happiness that a subject may expect from the institution or administration of a commonwealth but above all merely the right that is to be secured for each by means of it, which is the supreme principle for which all maxims having to do with a commonwealth must proceed and which is limited by no other principle. With respect to the former (happiness) no universally valid principle for laws can be given. (8:298)

Kant's claim, then, is that revolution can never be justified because it is motivated by a desire to secure laws, authorities, or institutions more efficient at producing happiness for the populace, and such a concern for happiness is never sufficient grounds for disrupting the rightful condition of the state. This argument is clearly insufficient, however, as we need not endorse Kant's apparent claim that revolutions are always motivated by a concern for happiness. If the citizens are instead motivated by a concern to correct for unjust laws, this argument would do nothing to explain why they act wrongly.[16] Kant needs the formal arguments from the "Doctrine of Right" to explain why even citizens motivated by justice cannot rebel against the state. For the remainder of the chapter, I will be focusing on the arguments from this later work, setting aside the happiness-based arguments from "On the Common Saying."

[16] See Guyer, p. 285, Williams p. 205.

Once we move on from the obvious limitations and failures of his earlier arguments, I think that the legal prohibitions Kant establishes in the "Doctrine of Right" are fundamentally consistent and even necessary. His arguments against a legal right to resist the state or rebel are ultimately rooted deeply in his foundational political and legal philosophy. Given the ways he has defined 'right,' it would not be possible to speak of citizens as having a right that is, in practice and in principle, unenforceable. Active resistance against the state's authority and even a passive refusal to obey a law would both threaten the possibility of a juridical state. As far as a legal right to resist the legislative authority itself, however, there is no way to make sense of how Kant could accommodate it.

There is a problem, though, when Kant attempts to draw a moral obligation out of this legal reality. As Hill notes, the *Metaphysics of Morals* goes to great lengths to divide ethical duties from juridical duties, and thus Kant needs further argumentation to bridge the gap between the two kinds of obligation.[17] While consistency entitles and even requires Kant to argue against a legal right to resist if there is no authority that can decide in one's favor, we need no such arbitration in order to consider an action morally permissible. After all, my actions can be moral or immoral prior to or outside of civil society, where no talk of 'rights' makes sense. If Kant wants to show that we have a moral obligation to obey the law – that civil disobedience, resistance, and rebellion are

[17] Hill, p. 290-291.

morally impermissible – then he must give us some reason beyond their mere illegality. Failing to do so is a clear conflation of legal and moral obligation.

To this objection, a defender of Kantian orthodoxy might respond that the failure to preserve a distinction between legal and moral obligation is no failure at all. Kant clearly holds that the law creates a moral duty where none existed before. The reason for this has to do with the origin of the law as a product of the general will. By giving our rational consent to whatever the legislative branch legislates, we essentially give the law to ourselves. In doing so, we place ourselves under an obligation to follow the law, no matter what our feelings about it might be. This obligation is legal, but it is also moral; any legal duty to obey the law would entail a moral duty to do the same.

Such a response, however, cannot truly answer the objection for one important reason: although any law does create a moral obligation, policies that require immoral action or to which citizens cannot rationally consent *cannot be laws*. Recall that this is one of the two limitations that Kant imposes on the legislative's ability to create laws. If the state attempted to pass a law instituting slavery, it would be one to which the citizens could not consent. As such, it could not be a product of their collective wills. Positing a moral obligation for the citizens to obey such a policy would entail creating a moral obligation for citizens to act contrary to what they or others could accept as moral agents. In effect, we would be morally required to act immorally. Even the authority of the sovereign cannot be sufficient to morally obligate an immoral action. This would be truly contradictory, and we would be left with no rational way to decide which obligation to obey.

This way of thinking runs counter to the view of Kant as a legal positivist that is defended by Jeremy Waldron. He argues that Kant should be understood as staking out a positivistic legal theory, where the legitimacy of the law is derived from the procedure by which it is produced, rather than some external moral standard.[18] He views Kant as refraining from basing the legitimacy of the law on some other, normative standard of evaluation in light of the fact of moral disagreement and the potential 'calamity' caused by such disagreement. Waldron explains that although Kant might be a moral objectivist, this does not rule out the possibility of individual's experiencing strong disagreements about morality and how to effectively secure happiness.[19] Furthermore, if steps are not taken to prevent this disagreement from occurring, the resulting disharmony can threaten the state itself, and thus destroy the rightful condition (and along with it, the possibility of property ownership). To negate this danger, Waldron sees Kant as resorting to a version of legal positivism. He describes his understanding of positivism as,

> the principle that an official should enforce the law even when it is in his confident opinion unjust, morally wrong, or misguided as a matter of policy. The enactment of the law in question is evidence of the existence of a view different from his own concerning the law's justice, morality, or desirability. In other words, the law's existence, together with the official's own opinion, indicates moral disagreement in the community. The official's failure to implement the law because he believes that it is unjust, or his decision to do some-thing other than what the law requires because he believes that action would be more just, is tantamount to abandoning the very idea of law.[20]

[18] Waldron, Jeremy. "Kant's Legal Positivism." *Harvard's Law Review*. Vol. 109, No. 70 (1996), p. 1541.
[19] Ibid., p. 1552.
[20] Ibid., p. 1539.

It is fairly clear how Waldron sees this description as applying to Kant's legal philosophy. The law is meant to take precedence over personal opinion, just as the moral law should trump our personal inclinations. This is a plausible account of how Kant envisions the interaction between the law and our own moral beliefs.

I agree with some of Waldron's points. Contrary to some interpretations,[21] Kant is not a natural law theorist. He clearly recognizes that the legitimacy of law as arising from way in which it was produced, rather than on its conformity with an independent standard of evaluation. Although he does confirm that our juridical duties are ethical, this should not be read as a claim that we have underlying moral reasons prior to the institution of the law. Rather, it is the fact of a positive law that gives us a corresponding ethical obligation. In Kant's eyes, any number of different laws, policies, and institutions can be legitimate, even though some of these might be far less efficient, stable, or fairly-balanced than others. Although Kant does have a progressive view of civil society (indicating an interest in seeing less-desirable laws be replaced by better laws), the inferior laws are still legitimate, provided they arise in the right way.

This last condition, however, is stronger than Waldron seems to acknowledge. The fact that a large range of policies are not appropriate subjects for law strikes me as a large difference from a purely positivistic picture. This limitation is based on Kant's underlying moral philosophy; the state cannot pass laws to which even one citizen could not rationally consent, for to do so would be to violate the respect owed to this

[21] See Mulholland, pp. 10-15

individual as a free moral being. It is for this reason that I claim Kant ought to be considered a 'constrained positivist.' While all laws that can exist are justified in positivistic ways, there are a wide range of policies that cannot be made into law for moral reasons. Returning to Kant's example of slavery, given that individual citizens cannot will themselves to be made into slaves, and the legislative authority of a state derives its power to create laws from the collective will of the citizens, it also lacks the power to will a law that reduces any citizen to a condition of servitude. Although the state might attempt to pass such a law and even enforce its execution, the state would be defending an illegitimate policy.[22]

If this is correct and such immoral policies do not acquire the status of law, then citizens are under no moral obligation to obey them. This alone does not, however, extend to citizens a legal right to resist the state's power, much less rebel against the state itself. All it shows is that such commands or decrees fail to morally obligate citizens. They can passively refuse to obey them without doing moral wrong. This refusal, however, does not extend beyond a right to civil disobedience. The citizens are not morally authorized to engage in resistance to the state's power in other ways, as this would involve violating the duty to obey other, legitimate laws. Further, the citizens are not morally entitled to engage in all-out revolution over one immoral policy or a small number of such policies. Likewise, if the citizens resist punishment for their moral

[22] NB: This does not mean, however, that the citizens are legally permitted to rebel against such a policy. Although it might, in fact, not be a law, there would be no one with the authority to make such a determination. As such, all the problems that prevent a right to rebellion would still apply.

refusal to obey, they would be acting impermissibly. At this point, then, citizens are merely authorized to engage in moral acts of civil disobedience aimed specifically at the problematic policies.

In some cases, however, the circumstances might be such so as to extend the moral authorization to disobey and resist further. If the state goes beyond merely attempting to pass and enforce a policy that cannot be law, enacting a wide range of illegitimate policies or radically expanding its own power, then the citizens might be morally permitted to engage in a more general strategy of resistance to the state's power.[23] There would still be no legal right to do so, but the general obligation that citizens have to follow the law might be eroded to the point that the state is propped up by powers to which the people could never rationally consent.

In other cases, there might even be a moral obligation to engage in this widespread resistance. We might consider the duty that all people have to contribute to the progress of humankind. Part of this progress is the development, maintenance, and protection of rightful conditions. If citizens belong to the kind of abusive state we have been considering, then might not revolution prove the appropriate way to contribute? Lewis White Beck warns against this line of thinking: our duty to promote the progress

[23] There is no clear, fixed line that, once the state crosses, the citizens go from being morally authorized to engage in civil disobedience to being morally authorized to engage in resistance. Given that one of the major differences between the two is that resistance, if legitimate, involves the moral permissibility of resisting arrest and punishment, it seems likely that this shift would often involve increasing abuses of the state's executive or coercive power. Still, I am not offering either necessary or sufficient conditions for the premise difference between these two types of disobedience.

of humankind is an imperfect duty, and is therefore secondary to the perfect duty that all citizens have to obey the law.[24]

While Beck is right to suggest that our imperfect duty to promote the progress of mankind cannot trump a perfect duty to obey the law, this still assumes that the policies in question are, in fact, laws. As I have argued throughout, such policies cannot meet the requirements Kant imposes on law; they cannot truly be passed by a legislature. As such, we can have no moral obligation to obey. We might, in fact, be obligated to resist either the particular law or even the state's authority on the grounds of our imperfect obligation.

Could this duty to resist a state's slide toward tyranny go so far as a moral authorization or even requirement to engage in revolutionary actions? Kant takes a hard line against this possibility, arguing that revolution necessarily results in anarchy. As Guyer describes his thinking, "The overthrow of an existing state, even if in the hope of greater justice and not merely greater happiness, can never be an immediate transition to a better-constituted state, but is always a reversion to a condition of lawlessness. From such anarchy a better state *might* arise, but then again it might not."[25] It would be contradictory, then, for us to revert to lawlessness under the motivation of our duty to promote juridical states.

This line of thinking only seems to work when we consider a state that is still functioning in a quasi-rightful manner, albeit badly. If we imagine that the state has

[24] Beck, p. 420.
[25] Guyer, p. 287.

descended to the point of actively violating the rights of the people with great regularity and efficiency, there may be good reason to think that whatever state arises from the anarchy will almost certainly be better than the one we currently inhabit. Although we could never be truly certain about this, the worse our present state is, the more likely it becomes that whatever comes next will be better. Whatever obligation we have to the state would have eroded long ago, and at this point the state's authorities would maintain their power through the sheer use of force, unconnected with any authorization arising from the general will. Although there could still be no legal right to rebel against a state, there could be a moral one if all other avenues of reform had been exhausted.

The moral permission to disobey the law is grounded in the impossibility of the state obligating its citizens to behave in immoral ways.[26] Even if I am not personally affected by a law, I still have the grounds to object to it if it requires such immorality of anyone else. There can never be a legal right to disobey or rebel, but there is a moral permission in these specific circumstances. As the state strays further from a rightful condition, greater forms of disobedience become morally permissible. If the state became

[26] In chapter five, I raised the question of whether any law that constrains freedom unnecessarily is unwillable and therefore not a policy that could be made into law. If this is the case, then any such law could be passively disobeyed. Based on the way Kant has set up the purpose of the state and the limits on what can become a law, this conclusion seems unavoidable. This would have the practical effect of morally licensing a large amount of disobedience, even if such disobedience was not legally permitted and could still be rightfully punished. One might think that this is a problem for Kant's entire framework and project. I will remain agnostic on this question; for the purposes of this project, trying to construct an alternative that could avoid this issue would require the revision of far too much of Kant's foundational commitments.

unjust enough, it might even be possible that the people could have an obligation to resist its policies and, finally, even rebel and replace it with a newer, better state.

7.2 Punishing Former Authorities

Even if civil disobedience or revolution is permissible – or, in certain rare cases, morally required – it would be by no means the preferred method of resolving conflict. An authority figure might be engaging in activity that exceeds the limits of her sanctioned power, but we ought to try to remove this figure through established, legal means. Provided such efforts to replace an authority figure are successful – or provided that, should they fail, the citizens resort to removing the abusive figure through some act of passive civil disobedience or active resistance – we are left with the question of what happens next. What is to be done with those who previously held, and misused, authority? Kant has a very explicit answer for this: nothing. Punishment against former legislative or executive authorities is, in his eyes, impermissible. He goes so far as to say that the greatest crime that a people are capable of is the execution of a former monarch after ousting him or her from power. In describing this, he concludes that such an execution is "a crime from which the people cannot be absolved, for it is as if the state commits suicide" (6:322).

Now, some historical considerations: the horror that Kant feels—and that he assumes we share—is directed toward the execution of monarchs by their people following a forced abdication. The execution of Charles I and Louis XVI during the

English and French revolutions, respectively, are his paradigmatic examples. The scenarios we are likely to encounter in the world today are likely to be different, and it is not clear how Kant would feel about a more contemporary case in which the former official is tried by an actual court, not executed as a spectacle. It is also not clear how the elimination of capital punishment as a penalty for former rulers would affect his stance. While death seems to be at the heart of his visceral reaction, he still gives us an argument of sorts to demonstrate why he opposes any punishment against former government authorities.

In the "Doctrine of Right," in a footnote, he writes the following, important sentence:

> The state never has the least right to punish him, the head of state, because of his previous administration, since everything he did, in his capacity as head of state, must be regarded as having been done in external conformity with rights, and he himself, as the source of the law, can do no wrong. (6:321)

Unlike his earlier statements regarding resistance to the power of a current authority, here Kant is speaking directly to the question of punishing past authority figures. Unfortunately, this passage is one of the clearest examples of Kant's lamentable tendency to resort to occasional obfuscation about the republican separation of powers that he has previously established. It is not clear whether he is referring to an executive figure or legislative figure. He claims that the state cannot punish former rulers, because anything that the previous ruler did "must be regarded as having been done in external conformity with rights," and that the previous authority is, herself, "the source of the law." These two powers that Kant describes as belonging to the former head of state

actually belong to both the executive and the legislative branches of government. Rather than respecting the separation of powers that he champions elsewhere, here he seems to be discussing the ruler of a state in which both legislative and executive functions are fulfilled by a single person or office.

This directly conflicts with what Kant says at 6:317: "So a people's sovereign (legislator) cannot also be its *ruler*, since the ruler is subject to the law and so is put under obligation through the law by *another*, namely the sovereign." In other words, a single figure could not be both the ruler whose actions are always in external conformity with right and the source of the law itself. Given that he begins the statement with the term 'head of state,' I think it is appropriate to conclude that he has in mind the executive authority and that the elision occurring here is the attribution of legislative sovereignty to the executive.

To understand what Kant means when he says that the executive always acts in external conformity to right, we need to look briefly on what he has to say on the subject of the executive's role in punishment. He holds that the head of state, as the chief executive, is immune from punishment; if there were anyone who could punish him, then he would not actually be the chief executive (a familiar, quasi-Hobbesian argument). On this basis, Kant holds that the executive is essentially free from all duties of right with respect to the citizens of the state. As Kant makes clear at 6:232, "Right and authorization to use coercion therefore mean one and the same thing." As no one is authorized to use coercion against the executive, no one has any rights against her; put another way, she is therefore free from juridical duties. Given this, it is also accurate to

say of the executive that she cannot do wrong. Since one acts 'rightly,' or at least in conformity with external right, until such a time as one violates a legal duty, and the executive technically has no legal duties, the executive always acts rightly.

The fact that the executive authority always acts rightly, however, does not make her or him the source of the law. As we saw above, the executive is also obligated by the sovereign legislative authority, which Kant argues must be distinct from the executive. Further, given the particularity of the executive and generality of the legislative, they could not both be fulfilled by the same figure or office. Indeed, it is even perfectly reasonable to speak of an executive's behavior as being counter to the law created by the legislative, even if such behavior is not 'wrong' in the sense that Kant employs in the "Doctrine of Right."

Once we have recognized this conflation, we can set it aside as an error that Kant ought not to have made. If we respect the republican separation of powers, however, is there any reason to think that the legislative or executive authorities of a state cannot be punished after they are no longer in office? Kant himself clearly wants to preserve this permanent immunity; in the passage I quoted at the outset of this chapter, Kant writes that the head of state cannot be punished

> on the pretext that he has abused his authority. Any attempt whatsoever at this is high treason, and whoever commits such treason must be punished by nothing less than death for attempting to destroy his fatherland. (6:320)

Even if we accept the impermissibility of punishing a current authority figure, there are several difficulties involved in Kant's efforts to secure for authority figures a permanent immunity against criminal prosecution and punishment. The most obvious

and striking of these is the possibility that such punishment might not occur after forcible revolution or regime change, but rather simply after a state authority figure has left office. Imagine a scenario in which the current executive arrests her predecessor for abusing his authority and has him tried by a court. The current executive has an essentially unlimited right to punish; nothing she does constitutes a legal wrong. Yet according to Kant, the past executive is meant to be immune from such prosecution. If we take seriously Kant's claim that any attempt at punishing a former ruler should be punished by death, we seem to arrive at the peculiar result that the current executive must be executed. This, of course, would directly contradict much of what Kant says elsewhere.[27]

This answer, however, might not be fully satisfying. To argue that former executive authorities can be punished by the current executive solely in virtue of the connection between the head of state and the dictates of external right runs the risk of perpetuating the very kind of potential for autocratic abuse that I argued against in the previous section. We can avoid this problem by considering an alternative possible reason for a legal and moral right to punish former authority figures.

This alternative makes use of both a negative and positive argument. On the negative side, it is not clear why the past protections enjoyed by the executive should continue once she is no longer the executive. The claim that the executive is

[27] In addition to the obvious contradiction of saying that the executive can do no wrong but should be put to death for executing a punishment against a citizen, he also makes it clear that even if a government comes into being in an illegitimate way, it still commands the obedience of the citizens (6:318-6:319). In light of this, it is hard to imagine how a newly elected executive would deserve death for punishing the old executive, who has left office and is once more a mere citizen.

unpunishable refers to the fact that there is no one in a position to carry out the punishment, not, as Kant suggests by way of conflation, that whatever the executive does is transformed into a legally valid action. Given that the executive has no coercively enforceable duties, she has no legal obligations and thus cannot do 'wrong.' In this sense, her actions are always right. This does not mean, however, that the executive cannot be responsible for breaking a law that was created by the legislative. Even if her office protects her from any penalty for such an action, she only retains this office for as long as the legislative sees fit. Once removed, she is a citizen who can be held accountable for her actions, just like any other.

Kant might argue that criminally charging a former authority for something that she did while in office is tantamount to criminizaling behavior after the fact. It was right when the executive acted, and so it seems unfair to punish her now that it would be wrong for her to do it. This parallels the long-standing objection that liberal thinkers have to *ex post facto* laws. While there is a superficial similarity between these two cases, there are underlying differences. In the case of *ex post facto* laws, the guilty citizen had no legal or moral obligation to refrain from engaging in the action when he or she committed it; the executive, on the other hand, has unenforceable duties to obey the law by virtue of her membership as a citizen of the state. It seems, then, that not only would an executive act rightly when punishing his predecessor, but that such punishment could be morally justifiable as well.

A legislator, on the other hand, cannot be said to always act in conformity with external right. If a legislator takes bribes and is removed from office, for instance, then it

seems difficult to see what possible reason Kant could give for withholding punishment. Recall that the legislative power Kant ascribes to his vague, unpunishable figure in the quotation from 6:322 is being the 'source of the law.' This, however, only applies to the legislative body as a whole; the individual malfeasance of a particular senator or member of parliament does not take on legal status. Even if there is a single legislative figure, his will only becomes law under certain circumstances. For instance, it must be general and universally willed. If he chooses to accept a bribe in a specific instance, this does not amount to the sovereign willing a law that bribes are allowable in all cases. While it might be difficult to punish a legislative figure for a law that he or she created, it does not seem hard to recognize the difference between what the legislator does *qua* legislator and what he does *qua* individual citizen.

The negative argument, then, shows there is no good reason to extend permanent immunity to prior authority figures. The positive argument, on the other hand, has to do with the justification for punishment itself. Interestingly, Kant's hesitance to punish the previous authorities of a state seems to run counter to his espoused retributivism. We might, perhaps, read him as thinking that the authorities deserve punishment, but that they are merely shielded from it by the formal structure of power with respect to their offices. If this were the case, however, why would such protection continue once the authorities no longer occupy the offices in question? It makes more sense to understand Kant in the manner that I have argued for throughout chapters four and five: namely, that Kant does not provide a means for authorizing the state to always – or possibly ever – act on moral desert. If the authorities can deserve to

be punished but escape for formal legal reasons, why should we think that the average citizen is punished for reasons of desert, rather than also for formal legal reasons?

In light of this, I think it worth our time to consider the case that Kantian protective deterrence could make for punishing former authorities. As we have seen, the reasons given by Kant for refraining from such punishment are not satisfactory, but this alone is not an argument in favor of punishing. If we reject retributivist reasons for punishing abusive authorities after their time in office, the mere fact that we *can* punish them does not tell us that we *should*. However, if we embrace a deterrent reading of Kant, of the sort that I have championed, it is not difficult to provide such an argument. Clearly there is great danger in executive and legislative misconduct or overreach. Such abuses are among the primary cause of the dissolution of republics into autocratic or anarchical states, to say nothing of the actual harm and loss of freedom that the citizens stand to suffer. It seems, then, that there would be ample reason to want to deter this kind of impropriety. If there is never any penalty for abusing one's powers while in office, it is hard to imagine from where the deterrent force would derive. Although we might optimistically hope that our leaders are driven by a strong moral commitment to duty, we cannot rely on this alone; to do so would violate the Kantian determination to construct a state in which even a race of devils could live rightfully (8:366).

Kant even has a built-in mechanism for determining that the punishment of a former authority figure will happen in the proper manner: the judiciary. Kant talks of the judiciary as a separate branch of government (6:318), but I have not included it as such given that the interpretive consensus is that Kant's judiciary is not, in fact,

independent. After all, the executive appoints all judges, and they serve the executive as agents (6:316-6:317). Given this, it seems more plausible to think of the judiciary as part of the executive branch of government.

Kant describes the role of judges quite eloquently: "A people judges itself through those of its fellow citizens whom it designates as its representatives for this by a free choice" (6:317). Like the laws created by legislative branch, the verdicts of the judiciary are meant to reflect the judgments of the people as a whole.[28] Although it is left vague in what manner a judge ought to reflect on a case so as to ensure that his decisions are appropriately general, the sentiment of a people judging collectively through judicial representatives is both a compelling image and one that is especially well-suited to the subject of passing sentence on a former authority figure.

Rightful punishment of such a figure could be permitted under Kant's political philosophy. His argument against it conflates the legislative and executive powers and relies upon a faulty assumption about a permanent immunity against prosecution. It is also motivated in large part by an image of ruler punishment that depicts a deposed figure being executed by mob rule. To avoid these issues, the citizens of a state must wait until 1) the abusive authority in question no longer occupies her former office and 2) there is an executive authority in power who has appointed or can appoint a judge to

[28] NB Although the judiciary reaches a verdict of guilty or not guilty, it is not the judiciary that decides upon the appropriate penalty to be imposed on a perpetrator who is found to be responsible. If the penalty is definitively articulated in the law – all murderers are to be executed, for instance – then we could say that the legislative is responsible to fixing the penalty. If, on the other hand, there is no determinate punishment called for by the law, then it is up to the executive to determine what penalty is appropriate. It is also worth noting that this division of powers is true both for Kant's orthodox retributivist position and the Kantian deterrence view I have defended.

determine the innocence or guilt of the former authority. If a former executive is being tried, then a replacement must be in place; if it is a former legislator, then there is already the appropriate executive and judicial apparatus necessary for a fair trial. As long as these conditions have been met, then we should consider Kant as capable of allowing for the punishment of past authorities by domestic courts.

7.3 International Punishment

Up to now, we have been focused on what rights, obligations, or powers the citizens of a state hold against the legislative or executive authorities of that state. It is also worth considering, however, what role Kant's cosmopolitanism plays in his thinking about resisting the power of state authorities. Specifically, I am interested in the whether the authorities of a sovereign nation can rightly be coercively constrained, tried, and punished by the force of another state. There are, I believe, two distinct reasons for extending our thinking to the level of international politics.

First, Kant himself has a clear interest in the application of his political thinking to cosmopolitanism. Although his considerations of the 'rights of nations' is a relatively minor aspect of the *Metaphysics of Morals*, his influential essay *Toward Perpetual Peace* provides a more complete picture of Kant's position on international and cosmopolitan law. Unlike the seminal political treatises of Hobbes, Locke, Rousseau, and others – which stopped at a description of the appropriate way for states to interact with one another – Kant considers the future development of the international community.

Although there is considerable scholarly dispute over the exact nature of the form of international relations for which he advocates, it is clear that he sees his contemporary status quo as a temporary step along the way toward achieving the highest good.

The second reason for exploring the question of international punishment has to do with the traditional categories of just war described by the proponents of natural law. At least as far back as Grotius, and continuing through Pufendorf, Locke, Vattel, and Burlamaqui, the natural law tradition has recognized punishment as an appropriate cause for war. Note that this is distinct from self-defense; the legitimacy of a punitive war means that any nation would be justified in initiating conflict with another that has participated in violations of natural law. Given the prominence of punitive wars in the *ius ad bellum* literature and Kant's unique foray into the genre with *Toward Perpetual Peace*, it is worth investigating whether he continues or breaks with the traditional categories.

Despite the significant historical precedent, though, the weight of evidence suggests that Kant would not support the international use of force for punitive reasons. Kant is not a natural law theorist, and so despite his affinity for Pufendorf, Achenwall, and others, we ought to be cautious about assuming that he continues to use their traditional categories of just wars. He indicates such a radical break in *Toward Perpetual Peace*, where he attempts to dispel the idea of a justified war and suggest steps that could be taken to the elimination of all war. This aim is partially at odds with Kant's statements about the rights of nations in "Doctrine of Right," in which he indicates that there are certain types of war that can be rightful, like self-defense (6:346). In this same

section, however, Kant declares that punitive wars are not among those to which nations can claim a right, so long as the nations exist in a state of nature together. His reasoning is analogous to the case for individuals: as I showed in chapter three, Kant does not recognize the institution of punishment as existing outside of civil society. Although there may be legitimate uses of violence (e.g., self-defense), such use of coercive force only becomes punishment when it occurs in accordance with established laws and recognized authorities with the power to enforce such a law. In situations in which a state of nature exists between nations, then, there can be no punitive war. The question, then, is whether Kant's thoughts on cosmopolitanism and his ideals for international relations give us any basis for thinking that states may not remain in the state of nature with respect to one another, thus opening up the possibility that punishment could exist at the international level.

In *Toward Perpetual Peace*, Kant lays out three 'definitive articles' that, if followed by all nations, would lay the groundwork for a permanent international peace. The first of these is that all states should be governed by republican constitutions. Given that I have been focusing exclusively on republican states throughout the chapter, this article will make no substantial change in the way we think of authority. The second definitive article, however, is very important for our present purposes: it states that "The right of nations shall be based on a *federalism* of free states" (8:354). As we will see, the nature of this federalism is hotly contested among Kant scholars, and it will play an important role in determining the rights that states have to punish abuses and crimes that occur in

others. Likewise, the third definitive article – "Cosmopolitan right shall be limited to conditions of universal *hospitality*" (8:357) – will feature significantly in our discussion.

There is a general consensus that Kant's call for a federation of free states is not intended as an endorsement of the establishment of a single world-state. Kant himself is relatively explicit about this; he writes, "here we have to consider the right of *nations* in relation to one another insofar as they comprise different states and are not to be fused into a single state" (8:354). His main reason for rejecting such a world-state is that it would require individual states to surrender their independence. While individual persons in a state of nature would be capable of joining together into a civil community, states that are analogously situated cannot bind themselves into a single political body. Doing so would be an act of the collective will of the people or community, and yet unlike the individuals who join together to form a nation, the community would be obliterated by the decision to join into a world-state. The old community would represent a transient intermediary; only the new community and the individuals would still remain. In light of all this, Kant does not think it possible for a political body to electively join a world-state and thus end its own existence.

It is tempting, then, to think of the federation of free states that Kant calls for as being comprised of totally independent nations. If this is the case, though, then it is hard to imagine in what sense any federation actually exists. Such a federation would have no authorization to pass laws, coercively enforce standards of any kind, or in any other way make its presence felt by the member states. The moment they choose to disregard the federation, they would be free to do so. Yet, Kant clearly thinks that this federation of

free states is a necessary step toward perpetual peace, and he explicitly links the gradual inclusion of all states into the federation and the development of a lasting international peace (8:356-8:357). In light of the role that the federation of free states is meant to play in the promotion of such a peace, joining it becomes a duty of some form.

Byrd and Hruschka point out that this obligation – which parallels the one that enjoins individuals to enter civil society and authorizes the use of force to compel those who refuse – is also one that contains the authorization to resort to the application of coercive force against those who would reject it.[29] In other words, a state can be compelled to join the federation of free states. This force is justified in light of the danger that proximity and interactions with a state that is not aligned with the federation could involve. As Byrd and Hruschka explain, "In a state of nature, the states cannot assert their rights through a 'proceeding' in court, because the state of nature is defined as a state without distributive justice, meaning for Kant there is no court with the coercive force needed to enforce its decisions."[30]

Such a reading seems to directly oppose what Kant lays out in the fifth 'preliminary article.' Here, he defends the view that "No state shall forcibly interfere in the constitution and government of another state" (8:346). Once again, there is a conflict between the independence of the state and the influence that can be exerted over it by its peers. In this case, however, the conflict is more easily resolved. We should understand

[29] Byrd, Sharon B. and Hruschka, Joachim. *Kant's Doctrine of Right: A Commentary*. Cambridge: Cambridge University Press, 2010, p. 194.
[30] Ibid.

Kant as holding the position that the coercion of states into the federation occurs as an act of self-defense. His prohibition against the reformation of the constitution or government of another nation, on the other hand, is aimed at aggressive wars, or those designed to bring about some result like religious conversion. While we are left with a possible question about the permissibility of altering a state's constitution if it is the only way to possibly bring it into compliance with the requirements of the federation of free states (i.e., a republican constitution), this can be set aside for the moment.

So a state might be justified in bringing another into the federation of free states if the other poses some security risk, on the grounds that members of the federation will have a greater chance of attain peaceful relations. Other forms of aggressive war, including punitive war, are not allowed. Further, wars among the members of the federation are prohibited; the whole point of the association is to promote peaceful relations. Even if they were not, the federation does not establish any kind of law or have the authority to enforce regulations, so punitive wars among the members of the federation also appear to be strictly impossible. So far, there does not appear to be any Kantian grounds for international criminal courts or other forms of non-domestic punishment.

When describing the prohibition against international interference with a nation's constitution, Kant does establish one important caveat. If the state in question should be engaged in a civil war, such that the international interference amounts to

aiding one side,[31] then such an intercession is permissible (8:346). Either the state has dropped into anarchy and there is no national sovereignty to be violated by such an action, or the legitimate government is at war with a powerful enemy that is no longer controlled by its authority. In the latter case, aiding the legitimate government in its efforts to subdue its enemies would not count as a violation of its independence or sovereignty (provided it has the permission or consent of the legitimate government to offer such assistance). In both cases, this kind of intervention could be carried out by members of the federation if they observe a civil war reaching the crisis point in one of their peers. This fact would allow for the institution of international criminal courts in one highly specific kind of case: namely, that in which the international community arrests an executive who is at war with the legitimate, legislative power of the people. While it is possible to give reasons for why it would be better for the rogue figure to be tried by her own community, it might be possible that the civil war has eliminated one or more of the necessary conditions for domestic punishment of a former authority I discussed above. In this case, the federation of free states could punish the figure without committing any violation of autonomy or independence.

Short of such an emergency, however, a Kantian right to intercede in the affairs of another nation with the interest of punishing its rulers for abusive violations of power would be dubious at best. The tyrannical use of power by some authority figure in another country is, essentially, not a problem for the federation of free states. As long as

[31] It would not be permissible, for instance, to interfere in a civil war with the aim of destabilizing the situation further.

this tyrannical figure maintains an orderly state, there would be no right to intercede. To ascribe such a right would tip the delicate balance of the federation of free states too far in the direction of a world-state.

This fact might seem dissatisfying, but I think we need not despair at the prospect. As I argued in the second section, the Kant's prohibition against punishing former state authorities is not one Kantians need necessarily endorse. There are good reasons for thinking that such punishment is both permissible and desirable for its deterrent effects. While this might not lead to an immediate remedy to the problem of a tyrannical ruler, it does allow for progress to be made in the long run. Finally, if the abuses become too great for a people to bear, their resistance and eventual revolution can claim moral – if not legal – justification.

Conclusion: The Future of Kant's Theory of Punishment

According to some estimates, there are between ten and eleven million people currently incarcerated across the world.[1] This represents an increase of 25-30% over the past fifteen years. The United States of America has one of the highest incarnation rates of any country, and when other legal sanctions – such as fines – are factored into the total, as many as two million American citizens will be punished by the criminal justice system each year—as many as 10,000 of whom may be innocent of the crimes of which they are convicted.[2] Legal punishment is a reality of many people's lives, and one that seems plagued by systemic racial, social, and economic injustices. Given the continued commission of crime, the high rate of recidivism, and the tragic punishment and even execution of those later exonerated, our criminal justice system seems deficient at achieving deterrence, rehabilitation, or even appropriate retribution. For all these reasons, there is a powerful and urgent need to seriously reexamine the traditional frameworks and theories that justify and guide the state's use of punishment.

Immanuel Kant's theory of punishment occupies a prominent place in this traditional canon. The orthodox reading of Kant's theory is elegant and straightforward: the only morally permissible justification for punishment is retribution for a previous

[1] All statistics about domestic and international punishment rates can be found in United States Bureau of Justice Statistics. "Prisoners in 2008." United States Department of Justice, 2009; and Walmsley, Roy. "World Prison Population List (tenth edition)." *International Centre for Prison Studies*. 2013.

[2] Huff, C. Ronald et al. *Convicted but Innocent: Wrongful Conviction and Public Policy*. Los Angeles: Sage Publications, 1996.

wrong, and the form of this punishment ought to resemble the crime as much as possible. He was among the first Western philosophers to ever give a secular account of retributivism as the justification for punishment, and consequently he has come down through history as the grandfather of retributive theory. His support for retributivism went largely unchallenged for the better part of two centuries, at least partly because it fit within an easy narrative. Consequentialist moral and political theories were well-known supporters of deterrence as the sole justification for punishment, and as the most significant philosophical alternative to consequentialism, Kant's practical philosophy seemed to be a natural place to find retributivism.

As Kant's political philosophy gained greater attention in recent years, this narrative came under serious scrutiny. The basis on which Kant grounds his retributivism has been questioned; in particular, the role of moral desert in his political philosophy seems problematic. In place of this retributivism, Kant's interpreters have suggested deterrence theories, rehabilitative theories, and mixed theories. Even some of those who at one time defended the traditional reading of Kant as a retributivist eventually conceded that his arguments do not successfully support a retributive theory. While many generic treatments of punishment still list Kant as foremost among the philosophical supporters of retributivism, a significant portion of the scholarship about Kant has moved on from the traditional reading.

I concur with these assessments. Kant's retributive theory of punishment has several insurmountable issues, and his practical philosophy is both capable of supporting and more consistent with alternatives that rely on deterrence for a

justification. Most fundamentally, Kant's conception of the state and its purpose is at odds with the idea that the state is authorized to punish based on moral desert. Even if we accept that Kant holds that doing wrong makes one morally deserving of suffering – itself a contentious claim – this does not show that the state is justified in bringing about this suffering. In order to connect the state's power to punish to moral desert, he would need to give an account of why the state has the power to respond to the moral desert of those who violate their juridical duties, as well as explain why the state is not authorized to punish – regardless of our moral desert – when we fail to satisfy our ethical duties.

Kant does have the resources available, however, to construct a theory with a deterrent justification. In particular, I have defended 'Kantian protective deterrence' – a mixed theory of punishment that incorporates retributive and rehabilitative elements. By separating out the liability, method, and amount of appropriate punishment from the justification offered for punishing, this theory is able to capture a number of the different interests and goals we have for punishment, all within a deterrent framework that is consistent with Kant's variety of liberal political theory. Kantian protective deterrence departs from other efforts to construct deterrent or mixed Kantian theories of punishment in its focus not on deterring threats to the state's authority, but rather deterring threats to the individual freedom of each and every citizen.

Kantian protective deterrence is not intended to accurately represent Kant's historical views. At the same time, it is not merely a theory of punishment loosely inspired by the practical philosophy of Kant. Rather, I have sought to establish an alternative theory that is grounded firmly in his most fundamental commitments. In

addition, I have aimed to preserve as many of Kant's original claim about punishment as possible; in the cases that this was not possible it was because such claims contradicted core tenets of his practical philosophy. As I have argued throughout the dissertation, this alternative not only avoids the hard question of the state's connection to desert, but it also is more harmonious with Kant's foundational practical philosophy in several key respects.

Removing Kant from the retributivist canon also changes the landscape of the contemporary debate about punishment. Without the ability to reflexively rely on deontology as a support, retributivists will need to be more explicit about the moral and political foundations of their view. In exploring the considerations that make retributivism untenable for Kant, we have also been exploring what it would take for any theory to provide a fully consistent account of a retributivist justification for punishment. Any retributivist working within the liberal tradition would face challenges in avoiding the same kinds of difficulties that I have described throughout the dissertation.

Going forward, there are profound practical implications of Kantian protective deterrence. Criminal justice – particular in the United States – is comprised of a patchwork of competing goals, interests, and ideologies, and this lack of theoretical consistency is reflected in the serious and systemic flaws in the execution of punishment. Adopting Kantian protective deterrence would require serious changes in the form that punishment takes, the kinds of laws and sanctions that are created, the way in which criminals are housed and treated, the length of sentences, and the goals associated with

punishing. In future work, I will explore the particular details of how our institutions and practices would shift in the event of this change. For now, I will simply say that truly adopting Kantian preventative deterrence would require laws that respect freedom, sanctions that are designed to deter crime, institutions that take individual responsibility seriously, and punitive practices that genuinely aim to rehabilitate the convicted.

The foundation for all these changes is Kant's guiding focus on the incomparable value of each individual, autonomous person. Kant's moral philosophy is an inspiring testament to the moral dignity of persons and the importance of freedom, and his political philosophy manages to represent these values while preserving the liberal principles of independence and tolerance. Kant deserves a theory of punishment that accurately reflects and compliments the strength and depth of his practical philosophy. Kantian protective deterrence is an effort to supply such a theory and to make Kant's philosophy a viable solution to the contemporary world's practical problems of punishment.

Bibliography

Allison, Henry. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press, 1990.

Altman, Matthew C. "Subjecting Ourselves to Capital Punishment: A Rejoinder to Kantian Retributivism." *Public Affairs Quarterly*, Vol. 19, No. 4 (Oct., 2005), pp. 247-264.

Ataner, Atilla. "Kant on Capital Punishment and Suicide." *Kant Studien*. Volume 97, Issue 4, Pages 452–482

Beccaria, Cesare. *On Crimes and Punishments and Other Writings*. Richard Bellamy, Ed. Cambridge: Cambridge University Press, 1995.

Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason*. Chicago: Chicago University Press, 1960.

---------. "Kant and the Right of Revolution." *Journal of the History of Ideas*, Vol. 32, No. 3 (Jul. - Sep., 1971), pp. 411-422

Beever, Allan. *Forgotten Justice: Forms of Justice in the History of Legal and Political Theory*. Oxford: Oxford University Press, 2013.

Berman, Mitchell. "Punishment and Justification." *Ethics*, Vol. 118, No. 2 (January 2008), pp. 258-290.

---------. "Two Kinds of Retributivism." *Philosophical Foundations of Criminal Law*. R.A. Duff and Stuart Green, eds. Oxford: Oxford University Press, 2011.

Bertram, Christopher. *Routledge Philosophy Guidebook to Rousseau and "The Social Contract."* London: Routledge, 2004.

Buchner, Edward Franklin. *The Educational Theory of Immanuel Kant*. New York: AMS Press, 1971.

Burlamaqui, Jean-Jacques. *Principles of Natural and Political Law*. Petter Korkman, Ed. Indianapolis: Liberty Fund, 2006.

Butler, Joseph. *Five Sermons*. Stephen L. Darwall, Ed. Indianapolis: Hackett Publishing Company, 1983.

Byrd, B. Sharon. "Kant's Theory of Punishment: Deterrence in Its Threat, Retribution in Its Execution." *Law and Philosophy*, Vol. 8, No. 2 (Aug., 1989), pp. 151-200

Byrd, B. Sharon and Hruschka, Joachim. *Kant's Doctrine of Right: A Commentary*. Cambridge: Cambridge University Press, 2010.

Cahill, Michael T. "Punishment Pluralism." *Retributivism: Essays on Theory and Policy*. Mark D. White ed., Oxford: Oxford University Press, 2011.

Cassirer, Ernst. *Kant's Life and Thought*. James Haden, trans. New Haven: Yale University Press, 1981.

Del Vecchio, Giorgio. *Justice: An Historical and Philosophical Essay*. Fred B Rothman & Co, 1982.

Dezhbakhsh, Hashem et al. "Does Capital Punishment Have a Deterrent Effect? New Evidence from Postmoratorium Panel Data." *American Law and Economics Review.* Vol. 5, No. 2 (2003), pp. 344-376.

Dolinko, David. "Retributivism, Consequentialism, and the Intrinsic Goodness of Punishment."

Ebbinghaus, Julius. *Die Strafen für Tötung eines Menschen nach Prinzipien einer Rechtsphilosophie der Freiheit*. Pan-Verlags-Gesellschaft., 1968.

---------. "The Law of Humanity and the Limits of State Power." *Philosophical Quarterly, Vol.* 3, No. 10, 1953:14-22.

Ehrlich, Isaac. "Capital Punishment and Deterrence: Some Further Thoughts and Additional Evidence." *Journal of Political Economy*. Vol. 85, No. 4 (August 1977), pp. 741-88.

Fleischacker, Samuel. "Kant's Theory of Punishment." *Essays on Kant's Political Philosophy*. Chicago: University of Chicago Press, 1992, pp. 191-212.

Fordyce, David. *The Elements of Moral Philosophy with A Brief Account of the Nature, Progress, and Origin of Philosophy*. Thomas D. Kennedy, Ed. Indianapolis: Liberty Fund, 2003.

Gil, Amanda et al. "Secondary Trauma Associated with State Executions: Testimony Regarding Execution Procedures." *Journal of Psychiatry and Law*. Vol. 34 (2006), pp. 25-36.

Gorr, Michael. "Toward a Theory of Coercion." *Canadian Journal of Philosophy*. Volume 16, Number 3, Sept. 1986, pp. 383-406

Gregor, Mary J. *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik der Sitten*. Oxford: Blackwell Publishing, 1963.

Grotius, Hugo. *The Law of War and Peace*. Richard Tuck, Ed. Indianapolis: Liberty Fund, 2005.

Guyer, Paul. *Kant*. 2nd ed. London: Routledge, 2014.

---------. *Kant on Freedom, Law, and Happiness*. Cambridge: Cambridge University Press, 2000.

---------. "Kant's Foundations for Liberalism." *Jahrbuch fuer Recht und Ethik*. Vol. 5 (1997): 121-40.

---------. "The Crooked Timber of Humankind" *Kant's Idea for a Universal History with a Cosmopolitan Aim*. Amelie Rorty and James Schmidt, eds. Cambridge: Cambridge University Press, 2009, pp. 129-149.

Haakonssen, Knud. *Natural Law and Moral Philosophy: From Grotius to the Scottish Enlightenment*. Cambridge: Cambridge University Press, 1996.

Hart, H. L. A. *Punishment and Responsibility: Essays in the Philosophy of Law.* 2nd ed. Oxford: Oxford University Press, 1970.

Hegel, Georg Wilhelm Fredrich. *Elements of the Philosophy of Right*. H.B. Nisbet, Trans. and Allen Wood, Ed. Cambridge: Cambridge University Press, 1991.

Herman, Barbara. *The Practice of Moral Judgment*. Cambridge: Harvard University Press, 1996.

Hill, Thomas E., Jr. "A Kantian Perspective on Political Violence." *Journal of Ethics.* Vol. 1, No. 2, pp. 105 – 140.

---------. "Kant on Punishment: A Coherent Mix of Deterrence and Retribution?" *Jahrbuch fuer Recht und Ethik: Annual Review of Law and Ethics* Vol. 5, (1997), pp. 291-314

---------."Kant on Wrongdoing, Desert, and Punishment." *Law and Philosophy*, Vol. 18, No. 4 (Jul., 1999), pp. 430.

---------. "Punishment, Conscience, and Moral Worth." *Southern Journal of Philosophy*. Vol. 36, No. S1 (1997). pp. 51-71.

---------. "Treating Criminals as Ends in Themselves." *Jahrbuch fuer Recht und Ethik*, Vol. 11 (2003), pp. 30-31.

---------. "Questions About Kant's Opposition to Revolution." *The Journal of Value Inquiry*. Volume 36, Numbers 2-3, pp. 283-298.

Hobbes, Thomas. *Leviathan*. J.A.C. Gaskin, Ed. Oxford: Oxford University Press, 1996.

Höffe, Otfried. *Categorical Principles of Law: A Counterpoint to Modernity*. Trans. Mark Migotti. University Park: Penn State Press, 2002.

---------. *Kant's Cosmopolitan Theory of Law and Peace*. Trans. Alexandra Newton. Cambridge: Cambridge University Press, 2006.

---------. *Political Justice*. Trans. JC Cohen. Oxford: Polity Press, 1994.

Holtman, Sarah. "Toward Social Reform: Kant's Penal Theory Reinterpreted," *Utilitas*, 9 (1997), pp. 3-21

---------. "Retributivism, Kant, and Civic Respect." *Retributivism: Essays on Theory and Policy*. Mark D. White, ed. Oxford: Oxford University Press, 2011.

Huff, C. Ronald et al. *Convicted but Innocent: Wrongful Conviction and Public Policy*. Los Angeles: Sage Publications, 1996.

Home, Henry. *Essays upon Several Subjects in Law*. 1732.

Hutcheson, Francis. *A System of Moral Philosophy*. Ed. Donald Winch. Indianapolis: Liberty Fund, 2006.

Kant, Immanuel. *The Cambridge Edition of the Works of Immanuel Kant*. Paul Guyer and Allen W. Wood, eds. Cambridge: Cambridge University Press, 1992--.

---------. *Education*. Ann Arbor: University of Michigan Press, 1960.

Kaufman, Whitley R. P. "The Mixed Theory of Punishment." *Honor and Revenge: A Theory of Punishment*. Dordrecht: Springer Publishing, 2013.

Kelly, J.M. *A Short History of Western Legal Theory*. Oxford: Oxford University Press, 1992.

Kersting, Wolfgang. "Kant's Concept of the State." *Essays on Kant's Political Philosophy*. Howard Williams, ed. Chicago: University of Chicago Press, 1992.

Locke, John. *The Second Treatise of Government*. Ed. C.B. Macpherson. Indianapolis: Hackett Publishing Company, 1980.

Ludwig, Bernd. *Kants Rechtslehre*. Hamburg: F. Meiner, 1988.

Martinich, A.P. *Hobbes*. New York: Routledge, 2005.

Merle, Jean-Christophe. "A Kantian Critique of Kant's Theory of Punishment." *Law and Philosophy*, Vol. 19, No. 3 (May, 2000), pp. 311-338.

---------. *German Idealism and the Concept of Punishment*. Cambridge: Cambridge University Press, 2009.

Mulholland, Leslie. *Kant's System of Rights*. New York: Columbia University Press, 1990.

Murphy, Jeffrie G. "Does Kant Have a Theory of Punishment?" *Columbia Law Review*, Vol. 87, No. 3 (Apr. 1987), p. 530.

---------. *Kant: The Philosophy of Right*. Macon: Mercer University Press, 1970.

---------. "Kant's Theory of Criminal Punishment." *Retribution, Justice, and Therapy: Essays in the Philosophy of Law* (Dordrecht, Holland: D. Reidel, 1979), pp. 82-92

---------. "Three Mistakes about Retributivism." *Analysis*. Vol. 31, No. 5 (Apr., 1971).

Nozick, Robert. "Coercion." *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. Sidney Morgenbesser, Patrick Suppes, and Morton White, eds. New York: St. Martin's Press, 1969.

Pallikkathayil, Japa. "The Possibility of Choice: Three Accounts of the Problem with Coercion." *Philosophers' Imprint*, Vol. 11, No. 16 (November 2011), pp. 1-20

Parrish, John M. and Tuckness, Alex S. "Kant and the Problem with Pardons." (March 31, 2010). Western Political Science Association 2010, Annual Meeting Paper.

Pettit, Philip. *A Theory of Freedom*. Oxford: Oxford University Press, 2001.

Potter, Nelson. "Kant on Punishment." *The Blackwell Guide to Kant's Ethics*. Ed. Thomas E. Hill, Jr. London: Blackwell Publishing, 2009.

Pufendorf, Samuel v. *On the Duty of Man and Citizen According to Natural Law*. James Tully, Ed. Cambridge: Cambridge University Press.

Radelet, Michael L. and Lacock, Traci L. "Do Executions Lower Homicide Rates?: The Views of Leading Criminologists." *The Journal of Criminal Law and Criminology*. Vol. 99, No. 2 (2009), pp. 48-508.

Rajan, Nalini. "Is There an Ethical Basis for Capital Punishment?" *Economic and Political Weekly.* Vol. 33, No. 13 (Mar. 26 - Apr. 3, 1998), pp. 701-704

Rawls, John. "The Justification of Civil Disobedience." *Collected Papers*. Cambridge: Harvard University Press, 1999. pp. 176-189.

Raz, Joseph. *The Morality of Freedom*. Oxford: Clarendon Press, 1986.

Reiss, H. S. "Kant and the Right of Rebellion." *Journal of the History of Ideas*. Vol. 17, No. 2, Apr., 1956

Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge: Harvard University Press, 2009.

Rousseau, Jean-Jacques. *Discourse on the Origin of Inequality*. Donald A. Cress, trans. *The Basic Political Writings*. Indianapolis: Hackett Publishing Company, 1987.

---------. *On the Social Contract*. Donald A. Cress, trans. *The Basic Political Writings*. Indianapolis: Hackett Publishing Company, 1987.

Scheid, Don. E. "Kant's Retributivism." *Ethics*, 93 (1983), pp. 262-282.

Shuster, Arthur. "Kant on the Role of the Retributive Outlook in Moral and Political Life." *The Review of Politics*, Vol. 73, No. 3 (Summer 2011), pp. 425-448

Silber, John "The Importance of the Highest Good in Kant's Ethics." *Ethics*. Vol. 73 (1962), pp. 179-197

Smith, Adam. *The Theory of Moral Sentiments*. Fourth Edition, 1790. D.D. Raphael and A.L. Macfie, Ed. Indianapolis: Liberty Fund, 1981.

Steiker, Carol S. "No, Capital Punishment is Not Morally Required: Deterrence, Deontology, and the Death Penalty."

Timmons, Mark. *Kant's Metaphysics of Morals: Interpretive Essays*. Oxford: Oxford University Press, 2002.

Tonry, Michael (ed.). *Retributivism Has a Past; Has It a Future?* Oxford: Oxford University Press, 2011.

Uleman, Jennifer. "External Freedom in Kant's 'Rechtslehre': Political, Metaphysical." *Philosophy and Phenomenological Research*, Vol. 68, No. 3, May 2004, pp. 578–601.

United States Bureau of Justice Statistics. "Prisoners in 2008." United States Department of Justice, 2009

Vattel, Emerich de. *The Law of Nations*. Béla Kapossy and Richard Whatmore, Ed. Indianapolis: Liberty Fund, 2008.

Waldron, Jeremy. "Kant's Legal Positivism." *Harvard's Law Review*. Vol. 109, No. 70 (1996). pp. 1535-1566.

Walmsley, Roy. "World Prison Population List (tenth edition)." *International Centre for Prison Studies*. 2013.

White, Mark D. *Retributivism: Essays on Theory and Policy*. Oxford: Oxford University Press, 2011.

Williams, Howard. *Kant's Political Philosophy*. New York: St. Martin's Press, 1983.

Wolff, Christian. Ius Naturae and Ius Gentium. 1749.

------Philosophia Moralis. 1753.

Wood, Allen W. *Kantian Ethics*. Cambridge: Cambridge University Press, 2008.

---------. *Kant's Ethical Thought*. Cambridge: Cambridge University Press, 1999.

---------. *Kant's Moral Religion*. Ithaca: Cornell University Press, 1970.

Yost, Benjamin S. "Kant's Justification of the Death Penalty Reconsidered." *Kantian Review*, Vol. 15, No. 2, 2010.

Zailbert, Leo. *Punishment and Retribution*. London: Ashgate Publishing Company, 2006.