

Copyright

by

Arthur Shuster

2010

The Dissertation Committee for Arthur Shuster
certifies that this is the approved version of the following dissertation:

Ancient and Modern Approaches to the Question of Punishment:

Hobbes, Kant and Plato

Committee:

Thomas Pangle, Supervisor

James Hankinson

Russell Muirhead

Lorraine Pangle

Devin Stauffer

Jeffrey Tulis

Ancient and Modern Approaches to the Question of Punishment:

Hobbes, Kant and Plato

by

Arthur Shuster, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2010

Acknowledgments

Anything of value in this dissertation I owe to my teachers at the University of Texas at Austin, whose guidance and support throughout my studies have been a true gift. I particularly wish to thank Thomas and Lorraine Pangle, as well as Devin Stauffer, for sharing with me their insights into the great questions of political philosophy, without which this study would not have been possible. I also thank Russell Muirhead for his valuable encouragement and sage advice.

To my parents, Boris and Isabella Shuster, I wish to express my deepest gratitude for their steadfast support throughout my education.

A. S., 2010.

Ancient and Modern Approaches to the Question of Punishment:

Hobbes, Kant and Plato

Arthur Shuster, Ph.D.

The University of Texas at Austin, 2010

Supervisor: Thomas Pangle

Abstract

The modern criminal justice system is experiencing what may be called a moral crisis brought about by a fundamental disagreement regarding the just and humane treatment of criminals and the purpose of punishment. This crisis has been addressed by contemporary scholarship without much success. The most serious defect of these scholarly attempts has been a failure to grasp how the apparently clashing aims of punishment—deterrence, retribution, and rehabilitation—relate to the fundamental principles of modern politics. Without this knowledge, it is impossible to begin to understand how these different penal aims may today be compatible and how incompatible, or even to appreciate what is at stake in each of them.

In order to gain a firmer grip on the problem, this dissertation returns to the original arguments for modern punishment by examining crucial moments in its theoretical development. In Hobbes, modern punishment theory attains its first and most

consistent articulation. Hobbes shows that the principles of modern politics limit the scope of justice to the protection of private freedom and property, and thus necessitate that deterrence should be the dominant aim of punishment. In his reaction against Hobbes, Kant affirms the importance of human dignity and argues that a penal system of pure deterrence would threaten the humanity of the criminal. Kant presents retribution as a more noble aim of punishment and tries to defend it on modern grounds, although he ultimately fails in this task.

In light of the aporetic conclusion of the examination of modern punishment theory, this dissertation turns to investigate the classical approach to the question of punishment as it is expressed in the proposal for humane penal reform in Plato's *Laws*. In the *Laws*, the highest aim of punishment, as the city understands it, is shown to be moral rehabilitation, although retribution and deterrence are also incorporated into the city's actual penal code as a concession to necessity and to the limitations of the *thumotic* civic outlook. The most humanizing feature of the penal reform proposal in the *Laws* is, however, its philosophical analysis of the nature of crime.

Table of Contents

Introduction: The Moral Crisis of Modern Criminal Justice.....1

Chapter 1: Hobbes’s Modern View of Crime.....19

Chapter 2: Hobbes’s Moral Theory of Punishment.....42

Chapter 3: Kant’s Effort to Rehabilitate Retribution.....68

Chapter 4: The Principles of Penal Reform in Plato’s *Laws V*.....106

Chapter 5: Plato’s Philosophical Criminology in *Laws IX*.....125

Conclusion.....165

Works Cited.....173

Vita.....179

Introduction: The Moral Crisis of Modern Criminal Justice

For the past four decades, criminal justice in the United States has been suffering from a crisis of conscience. One indication of this crisis is the U. S. Supreme Court's attempt to wrestle with the question of the morality of the death penalty in the 1970's. In 1972, the Court mandated a moratorium on the death penalty with its decision in *Furman v. Georgia*. The question before the Court was whether the imposition and carrying out of the death penalty is "cruel and unusual" and thus beyond the constitutional power of the state by virtue of the Eighth and Fourteenth Amendments of the U. S. Constitution.¹ In their respective opinions, the five justices who made up the majority of the Court agreed only on the narrow ground that Georgia's penal laws caused the death penalty to be imposed in an irregular and arbitrary manner, making it cruel and unusual, but could not agree on whether or not the death sentence was unconstitutional in principle. The justices' opinions reveal that they were well aware that whether or not the death penalty is in principle "cruel and unusual" depends on what makes punishment just or moral, and this, they saw, was precisely what they could not agree on.

Perhaps the most interesting of the opinions of the *Furman* court is that of Justice William Brennan. Brennan notes that the Eighth Amendment's prohibition of "cruel and unusual punishments" does not define the meaning of "cruel and unusual." If the Court is to apply the clause faithfully, it needs a principled method for determining which specific punishments fall within the clause's prohibition in a way that reflects the broad moral

¹ *Furman v. Georgia*, 408 U.S. 239 (1972).

commitments implicit in the clause itself and in the Constitution. But what exactly are those broad moral commitments? At bottom, Brennan believes that the “cruel and unusual” clause prohibits the infliction of punishments that are degrading to human dignity.² Does the death penalty deprive human beings of their dignity? Before addressing this question, Brennan illustrates his general point by giving the example of a different punishment which he believes does degrade the individual: expatriation. The punishment of expatriation is degrading, and thus cruel and unusual, because it involves “a denial by society of the individual’s existence as a member of the human community”; membership in a human community cannot be denied to a person without also depriving him of his humanity.³ Brennan does not consider the possibility that the punishment of expatriation may be justified by the dire threat a certain type of criminal may pose to society by remaining in its territory, even if it should be granted that such punishment is degrading. Nor does Brennan consider whether it isn’t possible that someone may deserve expatriation as a fitting punishment for a crime like treason. If a man betrays his country he is, as it were, denying its right to exist, and thereby might seem to lose his right to demand of his country that it should honor his right to exist.

Without considering these alternatives, Brennan goes on to conclude that the death penalty, too, degrades human beings and thus falls under the Eighth Amendment’s prohibition, since death is so physically and psychologically painful that it makes normal human function impossible in the time prior to and during the carrying out of the

² Ibid., 408 U.S. 270.

³ Ibid., 408 U.S. 273-4.

sentence.⁴ But, again, we are compelled to wonder whether the death penalty couldn't be justified by society's need for security if it were the case that the death penalty was indispensable to effective deterrence. As with the question of expatriation, we can also ask whether death isn't sometimes deserved by the gravity or inhumanity of the most heinous crimes. Has the worst criminal not already degraded himself to the point that all a conscientious society can do is to put him out of his misery?

Eventually, Brennan candidly acknowledges the existence of the serious objections we have just raised, and his admission is instructive:

From the beginning of our Nation, the punishment of death has stirred acute public controversy. Although pragmatic arguments for and against the punishment have been frequently advanced, this long standing and heated controversy cannot be explained solely as the result of difference over the practical wisdom of a particular government policy. At bottom, the battle has been waged on moral grounds. The country has debated whether a society for which the dignity of the individual is the supreme value can, without a fundamental inconsistency, follow the practice of deliberately putting some of its members to death. In the United States as in other nations in the world, the struggle about this punishment has been between ancient and deeply rooted beliefs in retribution, atonement or vengeance, on the one hand, and on the other, beliefs in the personal value and dignity of the common man that were born of the democratic movement of the eighteenth century, as well as beliefs in the scientific approach to an understanding of the motive forces of human conduct, which are the result of the growth of the sciences of

⁴ Ibid., 408 U.S. 286.

behavior during the nineteenth and twentieth centuries. It is this essentially moral conflict that forms the backdrop for the past changes in, and present operation of, our system of imposing death as a punishment for crime.⁵

Although throughout most of his opinion Brennan presents as uncontroversial his standard—human dignity—for distinguishing between just and unjust punishments, here he acknowledges the existence of a fundamental controversy over the true standard of just punishment, or over the end punishment ought to serve. If the proper aim of punishment were to safeguard human rights, as Brennan understands them, then the death penalty would be unjust. On the other hand, if the proper aim of punishment were either retribution or maximizing general security, then the death penalty may in fact be just.

While Brennan's opinion shows his awareness of the existence of the fundamental moral dispute between "beliefs in retribution, atonement or vengeance" and "beliefs in the personal value and dignity of the common man," it gives little indication of how this dispute might be decided, and on what grounds. Brennan's final position is that, historically, the death penalty has been used less and less in democratic societies (although it is still used in some) which to him is an indication that the fundamental moral dispute will eventually be decided in favor of "the value and dignity of the common man." In other words, Brennan appears to trust in historical progress, and specifically, in the progress of general moral opinion, as distinguished from scientific opinion, from ignorance to wisdom. However questionable this assumption may be, it

⁵ Ibid., 408 U.S. 296.

does not diminish the significance of Brennan’s admission that the fundamental moral disagreement about the standard of just punishment to which he points exists, that it has existed “from the beginning of our Nation,” and that both sides of this disagreement continue to have its adherents. The existence of this disagreement became still more obvious in the four years between *Furman v. Georgia* and *Gregg v. Georgia* (the case that lifted the moratorium on the death penalty), when Congress and more than 35 state legislatures reenacted the death penalty for one or more crimes. Partly in view of this legislative response to *Furman*, the plurality opinion in *Gregg* acknowledged that retribution is *one* of the morally acceptable purposes of punishment, since this appeared to be the view of a large portion of the country and of its representatives.⁶

Neither *Furman*, nor *Gregg*, nor to my knowledge any of the major Supreme Court’s Eighth Amendment decisions that followed have resolved the fundamental moral disagreement over the end of punishment that all sides admit exists in the country. This is not to say that the Court is necessarily equipped to decide such moral questions, or that it is always desirable that it should try to do so; judges must apply the law while at the same time being respectful of precedent and accepted usage. In this case, however, the law—i.e., the “cruel and unusual punishments” clause—does not provide any clear guidance, and so the moral dispute becomes politically urgent.

The political urgency of the question of the proper ends of punishment is not a recent development, as Justice Brennan hinted in his opinion. The history of punishment in the modern West has been marked by an enduring effort on the part of Western

⁶ *Gregg v. Georgia*, 428 U.S. 183-87 (1976).

peoples, their statesmen, and their intellectual leaders to humanize their penal institutions. In the United States in particular, an ambitious penal reform movement began around the time of the founding, generating lively debate about both policy and principle in assemblies, political pamphlets, and scholarly works.⁷ Most of the reformers agreed that American penal institutions needed to be modernized in light of the discoveries of modern medicine and the burgeoning fields of psychology and criminology. Above all, they were enthusiastic about the newest technical innovation: the penitentiary.⁸

Yet the goals of the reformers were not always the same. Some of the American penal reformers, like Thomas Jefferson of Virginia and William Bradford of Pennsylvania, were followers of the great Enlightenment philosophers of the seventeenth and eighteenth centuries, and especially of Montesquieu and Beccaria, who argued that punishment should be dedicated to securing personal freedom and property—i.e., that it should aim at the prevention of crime.⁹ Effective deterrence of crime, they argued, depends much more on the certainty and regularity of punishments than on their severity. Consequently, these thinkers and their American disciples favored milder modes of punishment, such as imprisonment and physical labor (as long as these were applied with consistency), over the death penalty and other corporal punishments. Other reformers, like Benjamin Rush, were deeply influenced by Christianity and looked for guidance from the Scriptures regarding the moral treatment of criminals. This second group of

⁷ For an account of the political context of these debates, see Ronald J. Pestritto's *Founding the Criminal Law: Punishment and Political Thought in the Origins of America*, chs. 1-3.

⁸ See Adam J. Hirsch, *The Rise of the Penitentiary: Prisons and Punishment in Early America*; David J. Rothman, *The Discovery of the Asylum*.

⁹ See Cesare Beccaria, *Of Crimes and Punishments*, chs. 5, 15, 19, 20; Montesquieu, *Spirit of the Laws*, bk. 6, chs. 9 and 16. Cf. Thomas Jefferson, *The Complete Jefferson*, 61; William Bradford, *An Enquiry How Far the Punishment of Death Is Necessary In Pennsylvania*, 3.

reformers were ultimately more concerned with the salvation of the criminal's soul and advocated for the adoption of a penitentiary system because they thought that by separating prisoners and occupying their time with physical labor the penitentiary created the conditions in which prisoners would be more inclined to repent of their crimes and turn to God.¹⁰ Thus, despite its uniformity of means, classic penal reform had been driven by heterogeneous influences from the very beginning.

In more recent times, the punishment debate has been joined and taken in new directions by critics of the modern criminal justice system. In 1975, the Frenchman Michel Foucault published an influential work, *Discipline and Punish*, in which he gives an alternative account of modern punishment to the one to which most liberals subscribe. According to the conventional view, Foucault argues, modern penal institutions and practices embody the principles of humanity and rationality and are fully transparent products of popular choice and consent. In reality, however, these practices are autonomous forces of individual subjection and domination, the products of slow and enduring historical accretions rather than of rational choice. While many of the book's most important conclusions are highly controversial and its scathing criticism of the treatment of criminals by liberal societies most unflattering, this work nevertheless contains several thoughtful insights which call forth serious reflection about aspects of modern penal justice commonly taken for granted.

According to liberals, punishment ought to be used to benefit society as a whole and criminals themselves through deterrence and rehabilitation, and not for making

¹⁰ See Benjamin Rush, *A Plan for the Punishment of Crime: Two Essays*, 6-7, 12.

criminals suffer or for subjugating them to the will of the law-abiding part of society. Liberals agree that incarceration is the mildest and yet most effective means of serving this moral end since it restricts the freedom of the individual instead of inflicting physical pain or public humiliation, as did the scaffold, the gallows, and the pillory.¹¹ Yet Foucault claims to explode this myth of the humanitarian origins of the modern penal system. In the bulk of the book, Foucault takes pains to show how, despite its pretensions to mildness and benevolence, modern carceral punishment is in fact much more insidious and ultimately more oppressive than the premodern forms of punishment, despite the barbarity of the latter. In the first place, he denies the distinction between corporal and non-corporal punishments, claiming that, in essence, all are of the former sort.¹² But Foucault goes much further than this. Following what appears to be an essentially Marxist analysis, Foucault interprets the apparently humanized and enlightened features of modern punishment as, at bottom, calculated means of perpetuating a certain kind of political economy favorable to the ascendant bourgeois class. The apparently benevolent substitution of incarceration and hard labor for corporal punishment—ostensibly to encourage moral reform and to instill the virtuous habits of industry and discipline—is in fact only a mechanism for subjugating the criminal’s body and mind, making him more productive as a kind of slave worker for bourgeois society.¹³ Even the modern conception of the soul as a rights-bearing subject functions essentially as “the prison of the body”—

¹¹ *Discipline and Punish*, 11.

¹² *Ibid.*, 16.

¹³ *Ibid.*, 26, 79, 91-92.

i.e., as one of the mechanisms of domination.¹⁴ Finally, what is perhaps the most insidious feature of the modern penitentiary system, according to Foucault, is that far from decreasing crime and recidivism it actually multiplies the number of criminals and delinquents, thus fueling itself and creating a subjugated underclass which law-abiding society can continue to exploit.¹⁵

Although the modern penitentiary has in our time become a staple of the penal system, its failure to produce genuine moral reform, and its various problems of recidivism, maladministration, overcrowding, and internal violence are well known and justly criticized.¹⁶ Yet Foucault goes too far in characterizing these problems as integral to modern punishment, as such, rather than as contradictions of its principles. Foucault summarily dismisses the professed humanitarian aims of the liberal penal reformers as disingenuous “lyricism” intended to mask the cold truth of underlying self-serving calculation, and thus does not give the reformers a fair hearing.¹⁷ Nevertheless, despite his exaggerations, Foucault’s basic point deserves serious consideration. Stated more cautiously, Foucault prompts us to wonder whether the penal reformers were right to assume that social utility on the one hand, and the moral treatment of criminals on the other, are two ends that could be achieved equally well by the same means. Have past and present generations of liberal penal reformers been too optimistic and too naïve in

¹⁴ Ibid., 30.

¹⁵ Ibid., 30, 264-272.

¹⁶ On early doubts concerning the penitentiary’s capacity to produce moral reform, see Gustave de Beaumont and Alexis de Tocqueville, *On the Penitentiary System in the United States and its Application in France*, 80, 90. For a discussion of the penitentiary’s other dysfunctions, see Rothman, *The Discovery of the Asylum*, ch. 10.

¹⁷ Foucault, *Discipline and Punish*, 91.

thinking that the moral treatment of criminals does not require any sacrifice of security on society's part?

Contemporary scholarship has made some attempts to address questions about the morality of modern punishment raised by Justice Brennan's opinion and by Foucault's book. One important recent development in punishment scholarship is the resurgence of interest in retribution. Critics of the classic liberal approach to punishment—and particularly of its utilitarian forms—have observed that while this approach may have benevolent intentions, its logic dictates the use of means that would have an alarmingly „paternalistic' tendency.¹⁸ The classic approach relies on an understanding of crime as analogous to disease, and of criminals as a dangerous class of ill human beings. From this perspective, the scientific expertise of prison administrators entitles them—not judges—to dictate the kind and duration of treatment offenders should receive. Since offenders may be pathologically incapable of making the right choices, they may not know what is best for them and may resist attempts to cure them; in such cases prison administrators would be entitled to disregard the autonomy of inmates, and perhaps even some of their rights. Moreover, guilt, innocence, and desert would lose their meaning according to this view.

These consequences of the classic approach are unacceptable to its critics, and so it is suggested that punishment ought to be concerned not merely with crime prevention, but also, and perhaps even primarily, with retribution. Unlike deterrence and

¹⁸ See Herbert Morris, "Persons and Punishment"; C. S. Lewis, "The Humanitarian Theory of Punishment"; Jeffrie G. Murphy, "Marxism and Retribution," 242-43; Joel Feinberg, *Doing and Deserving*, ch. 5. The position against which these critics are reacting is well expressed in S. I. Benn, "An Approach to the Problems of Punishment" (see especially 331-34).

rehabilitation, retribution aims at proportionality between the gravity of the crime and the severity of the punishment and takes into account the aggravating and extenuating effects of moral desert. In this way, it limits the coercion to which the offender may be held liable to what he deserves and no more, regardless of how much benefit may accrue to society—or to the individual himself, according to medico-criminological opinion—from an increase in duration or severity of punishment. Further, retribution seems to satisfy the principle of fairness, according to which the equally distributed burdens of civil society must be borne by each individual so that the benefits of civil society may be enjoyed by all: whenever one person refuses to bear his share of the burden by violating the law, retribution seems to be called for in order to reestablish the equilibrium.¹⁹ Some retributivists have also argued for retribution on the ground that it performs the important function of expressing society's collective view of what is right and wrong, and that penalties should be designed with the aim of communicating this view, even if this should diminish somewhat their deterrent or rehabilitative effects.²⁰

The critics of the classic approach were, in turn, challenged to explain and defend retribution independently, on its own terms, which proved to be a difficult task.²¹ Most importantly, however retribution had been defined by its advocates—as protection from disproportionate penalties, or as equalization of civic burdens, or as pedagogical expression of moral principles—it ultimately seemed to aim at safeguarding individual

¹⁹ See Richard Dagger, “Playing Fair with Punishment.” Morris, *op. cit.*, also suggests, although only briefly, that retribution may ultimately be grounded in the principle of fairness.

²⁰ Joel Feinberg's *Doing and Deserving* (esp. ch. 5) is the classic account of the ‚expressivist’ conception of retribution. See also Igor Primoratz, “Punishment as Language.”

²¹ For a review of the objections made against the various types of retributivism and the rebuttals to those objections, see David Boonin's *The Problem of Punishment*, ch. 3.

freedom. Whether a stable civil order requires retribution as a pedagogical tool to instill the principles of a free society or to equalize its inevitable burdens, or whether the principle of retribution sets up a moral boundary protecting the individual from being dominated by the rest of society, the ultimate end of punishment amounts to the same as what it was for classic penal theory. Retribution, so understood, seems really to be concerned with correcting the means used by the classic penal reformers, and not with a wholesale rejection of their project.²²

The way seemed to be clear, then, for attempting a reconciliation of the apparently opposed aims of punishment. The most famous such attempt was made by John Rawls and H. L. A. Hart who combined the retributive and utilitarian principles into a hybrid theory which has been called “mixed” or “partial” retributivism. In separate essays, Rawls and Hart make essentially the same argument.²³ The problem, as they see it, is that both utility and retribution make a legitimate moral claim on us, and yet appear to be heterogeneous and even mutually exclusive principles. This had led some to conclude that either one or the other principle must be true, but not both. Against this view, Rawls and Hart argue that there is a distinction to be made between principles that justify a practice and principles that justify particular actions falling under that practice, and that the principles of utility and retribution fall into these two different categories. The principle of utility answers the question of the fundamental justification of punishment as

²² See Benn, “An Approach to the Problems of Punishment,” 326-28.

²³ Rawls, “Two Concepts of Rules”; H. L. A. Hart, “Prolegomenon to the Principles of Punishment.” The difference between the two is that while Rawls tries to defend the principle of retribution as a rule within a rule-utilitarian framework, Hart defends retribution as independently grounded in the principle of fairness. In his later *A Theory of Justice*, Rawls adopts the principle of justice as fairness and explicitly follows Hart’s view of punishment. See *A Theory of Justice*, 211-12 and 276-77.

a practice or institution, while retribution justifies and regulates the administration of punishment in particular cases. To illustrate this argument, Rawls suggests that we look at each of the two questions as being addressed to a different public official: the question of general justification is addressed to the legislator, who looks to the future good, while the question of particular application is addressed to the judge, who looks backwards in time to the culpability or innocence of the accused.²⁴ Hart does a slightly better job than Rawls of showing why he thinks there really are two questions being asked. He draws attention to the confused view of actual political actors, e.g., a British Member of Parliament, who will appeal at different times to different principles to justify his stance on punishment. On each occasion, the official assumes that only one single principle justifies punishment as a whole and thus contradicts himself, failing to notice that at different times he intends to respond to different questions of justification about different aspects of the institution in question—about its general purpose and about its case by case application, respectively.²⁵ (Hart calls this political actors’ “over-simplifying tendency.”)

The Rawlsian-Hartian solution to the apparent tension between retribution and utility appears to combine the best of both. In keeping with the classic liberal approach, punishment is understood to aim at the protection of personal freedom and property while forestalling the paternalistic excesses of that approach—as identified by its critics—through the application of the principle of retribution in particular cases, and thus ensuring that criminals are punished only to the extent that they deserve and do not become subject to the absolute authority of penal administrators. This attractive solution

²⁴ “Two Concepts of Rules,” 5-6.

²⁵ “Prolegomenon to the Principles of Punishment,” 1-4, 10-11.

became influential for a time, but eventually it, too, came under attack. To its critics, it appeared that the Rawlsian-Hartian notion of “mixed” retributivism was too artificial, drawing convenient but spurious distinctions between questions of general justification and questions of particular application.²⁶ On the contrary, the critics argued, it is precisely with the *general justification* of punishment that retributivists are concerned.²⁷ As one critic argues, the fundamental problem that liberals are unable to escape is that their attempt to justify punishment entirely in terms of deterrence makes punishment a mere tax on crime, and this disagrees with our basic moral sense that there is an important difference between taxes and punishments.²⁸ This problem haunts liberals even when they attempt to “mix” retributive features with an essentially classic-liberal framework. The most serious implication of this criticism is that it becomes entirely possible that a truly satisfying theory of punishment—one that addresses all of our basic moral concerns—would have to be grounded in non-liberal principles.²⁹

Modern criminal justice has been challenged by yet another group of critics. During the past thirty years, a “restorative justice” movement has emerged in the United States and around the world. The basic objection of this movement to traditional criminal justice is that the modern state’s monopoly over the administration of justice deprives local communities of their autonomy; excessively focuses on the individual; and

²⁶ See J. Angelo Corlett, “Making Sense of Retributivism,” 81-83. See also Boonin, *The Problem of Punishment*, 63-77, for a critique of Rawls’s rule-utilitarian version of the argument.

²⁷ J. Angelo Corlett, “Making Sense of Retributivism,” 82-83; Stanley C. Brubaker, “Can Liberals Punish?” 822-25 and 827-31.

²⁸ Brubaker, “Can Liberals Punish?” 829-30. On the distinctions we commonly draw between punishment and other kinds of coercive measures, see Feinberg, *Doing and Deserving*, ch. 5.

²⁹ See Ronald Terchek and Stanley Brubaker, “Punishing Liberals or Rehabilitating Liberalism?” 1315-16; Brubaker, “Can Liberals Punish?” 833.

exaggerates the antagonism between the offender and his community.³⁰ Restorative justice advocates propose to return some of the control over the administration of criminal justice back to local communities by introducing community conferences into the criminal justice process. Such conferences would apply justice through “reintegrative shaming” rather than through traditional punishment, thereby healing frayed social bonds and reviving a richer sense of community.³¹ The restorative justice movement may thus be said to embrace one of the goals of the classic liberal approach to crime—rehabilitation—while rejecting its means—the state penitentiary.³² This movement also seems to appeal, although in a very distant way, to the ideal of classical politics, according to which the individual belongs to his community. While the restorative justice proposal is certainly attractive, how the empowerment of communities and the use of “reintegrative shaming” might be combined with liberalism’s commitment to the rights and freedom of the individual remains an open question.

It should already be evident from this selective overview of the punishment debate that there exists some awareness that at least a few of the important problems encountered with the liberal approach to punishment may not lend themselves to solution within a liberal theoretical framework. I believe this is in fact the case, and the aim of this study is to show which problems are intractable for liberalism and why. I do so by looking at definitive moments in the historical development of liberal penal thought,

³⁰ John Braithwaite, *Restorative Justice and Responsive Regulation*, 11 ff., 45, 74, 100-3, 129. See also Paul Hahn, *Emerging Criminal Justice: Three Pillars for a Proactive Justice System*; Nils Christie, “Conflicts as Property.”

³¹ Braithwaite, *Restorative Justice and Responsive Regulation*, 80-88, 129-32.

³² Braithwaite, *Restorative Justice and Responsive Regulation*, 95 ff.

paying special attention to the crucial impact of Thomas Hobbes and Immanuel Kant. In taking a historical approach my purpose is not to provide a genealogy of liberal penal theory, but to show more clearly the fundamental arguments that still shape it today in their native form, before they became truisms. As I will show, the character of liberal penal theory is dynamic, by which I mean that the moral dilemmas with which we are now wrestling were already present in liberal penal thought from the very beginning and were, in fact, the fundamental cause of its subsequent development.

The classic liberal approach, as we shall see, is fundamentally informed by Hobbes, and it is Hobbes's authoritative account of the genesis of the modern right to punish from the principle of self-preservation that most clearly shows that deterrence is its only rightful aim, and that retribution and rehabilitation, as traditionally understood, are necessarily ruled out. Hobbes's arguments make clear that any attempt to reinstate retribution or rehabilitation will either contradict modern principles, or it will result in the dilution of the original meaning of retribution and rehabilitation and their assimilation to deterrence. Thus, if I can substantiate this claim, it would help to explain the difficulty that contemporary retributivists have had in justifying retribution on its own terms, as an independent aim of punishment. In Kant, we will see the original criticism of the familiar modern consequentialist penal system from an essentially modern perspective, and the clearest articulation of the original meaning of retribution by any modern thinker. Kant shows that, in its original sense, retributive justice does *not primarily* mean either protection from disproportionate punishment, or restoration of the fair balance of civic burdens, or the expression of society's collective moral understanding. Kant's impressive

attentiveness to moral psychology leads him to conclude that the deepest meaning of retribution is the distribution of happiness in accordance with moral goodness. Through his failed attempt to reestablish retribution on modern principles, however, Kant also proves, despite himself, that Hobbes's original analysis is correct—which is not to say that it is also morally satisfying.

By recapturing the original dialectic of modern penal thought, we will be in a better position to understand its fundamental limitations. By the end of Chapter Three, I hope to have identified these limitations and thereby to have established the aporetic status of the question of punishment in the *modern* context. That is, I hope to have shown that modern political thought ultimately fails to satisfy all our moral concerns with respect to crime and punishment. I will take this conclusion as warrant to turn, in Chapters Four and Five, to examine the approach to the question of punishment taken by classical political philosophy, as exemplified in Book IX of Plato's *Laws*, in the hope of shedding further light on our present perplexities by looking at the question of the morality of punishment from a perspective that had been one of the major alternatives to the modern one, and which the latter had consciously opposed.

The modern reader may not expect someone like Plato to share many of his serious political concerns—with regard to punishment or any other sphere of political life—given the differences between Plato's society and his own. And yet, once one has invested some time in engaging with the Platonic dialogues, one is surprised at how familiar are their themes, and how fresh and direct is the manner in which they are treated. It is difficult for a study that seeks to compare ancient and modern political thinkers to

avoid touching on some very large issues, to which it would be difficult to do full justice given the limited aim of the present study. Nevertheless, given the vexed state of the current punishment debate, I believe that to see the modern approach to the problem of punishment in light of the classical view is of great value, and would repay our additional efforts.

Chapter 1: Hobbes's Modern View of Crime

Punishment is an important topic for modern political theory because the limits of a government's right to punish serve to define the boundary separating arbitrary violence from the rightful use of coercion, and thus tyranny from legitimate government. If wisdom or virtue alone is an insufficient claim to rule, then men would have good reason to be wary of government having unlimited discretion to use its might as it sees fit for advancing the common good. This suspicion would to some extent be assuaged if the government's right to use force against its able-minded citizens (in all cases except national emergencies) were circumscribed within the limits of a normative theory of punishment. For example, before being liable to punishment, citizens must have been found guilty of some crime; this crime and the penalty attached to it must have been established by law; the punishment could not be disproportionate to the gravity of the crime; and so on. Any use of force by a government against its able-minded citizens outside of these normative restrictions would be a violation of its rightful authority. Hobbes believed that other political thinkers in the modern tradition had not done enough to clearly fix this boundary and attempted to remedy the problem himself. Thus, although Hobbes is best known for defending absolute government, he is in fact also the author of the first moral theory of punishment in the modern tradition.¹

¹ The fundamental importance of Hobbes for modern punishment theory is acknowledged by Cattaneo and Norrie. See Mario Cattaneo, "Hobbes's Theory of Punishment"; Alan Norrie, "Thomas Hobbes and the Philosophy of Punishment."

To appreciate Hobbes's concern, we might glance at the teachings of his predecessor Machiavelli and his contemporary Spinoza. In their attempt to articulate the modern project of man's self-liberation, both Machiavelli and Spinoza had taught that the right to rule ultimately rests in superior force, and that this force is antecedently bound by no law—whether natural or divine. This basic principle informs Machiavelli's infamous advice to princes about "cruelty well-used." Machiavelli recommended that, in order to secure their own rule over their subjects, princes should use crimes committed by their subjects as opportunities or pretexts for a show of overwhelming force. Punishments *must* be excessive, cruel, and bloody, Machiavelli argued, because human beings can be ruled effectively only when they are overawed by superior force.² To rule wisely, according to Machiavelli, is to rule without moral restraint.

Hobbes believed that this teaching was much too bold and too alarming ever to be accepted as public doctrine, even though he agreed with Machiavelli that overwhelming force is necessary for keeping men moderate and in check. However prudent it might be for an entire society to accept absolute government, its individual members would be too proud and suspicious to endure for long a subjection they believed to be slavish. While agreeing with Machiavelli and Spinoza about the urgency of providing theoretical justification for a man-made power capable of dominating all political, moral, and religious disagreements, Hobbes believed it was necessary to go beyond their harsh and morally unpalatable teachings to show how the subjection of naturally free beings to their equals could be morally justifiable, and even noble, and hence not slavish. For this reason

² Niccolo Machiavelli, *The Prince*, ch. vii.

it also became important for Hobbes to teach not only princes, but also their subjects and citizens.

Because Hobbes was the first to address this uniquely modern problem, he was acutely aware that any morally satisfying account of punishment would at the same time have to avoid restricting absolute sovereign power and frustrating the fundamental aim for which modern government is established—namely, the effective protection of individuals' lives and property and the maintenance of a stable civil order. Hobbes's awareness of the need to reconcile the purpose of punishment with its morally defensible administration is one of the reasons why it is important for us, in our present predicament, to return to his thought. Hobbes's solution to this problem was to argue that while the sovereign may not be limited by any other power, he remains bound by the "laws of nature"—dictates of reason forbidding the sovereign to punish his subjects under specified conditions (e.g., if they are innocent). Ultimately, as we shall see, Hobbes only partly succeeded in his task, and the extent to which he failed is the second reason why we must turn to his writings on punishment. For his failure determined the course of the subsequent development of modern punishment theory by Beccaria, and then later by Kant.

One more point must be made before we embark on our examination of Hobbes's view of punishment. The fact that Hobbes's political teaching is addressed not just to princes and philosophers but also to a general audience is of the greatest significance. Hobbes's goal was not only to communicate his philosophy to those who could understand it, but also to initiate a vast cultural and political revolution by having his

teaching popularized (*Leviathan* xxxi.41).³ As a writer, he was acutely aware of the intellectual and psychological differences among the various types of men, and above all the difference in their aptitude for scientific learning. Hobbes saw that most men did not set down clear definitions or use terms consistently in their reasoning (v.17-18, 21). It would have been foolishly hopeful for Hobbes to expect his radical critique of the established tradition to be accepted for what it was: a set of deductions necessarily following from clear definitions, either to be accepted or rejected along with those definitions.⁴ The best that Hobbes could have hoped for was that the majority of his readers would accept his new political principles as authoritative dogma.

More crucially, however, the traditional doctrines regarding justice and injustice that Hobbes had intended to supplant appealed to, and gained their strength from, men's passionate hopes.⁵ In order to put into effect his vast cultural and political reforms, Hobbes had to write at the same time to two different classes of men: those who could understand his radically new vision as supported by reason and who would work to establish it as the authoritative public dogma (cf. again xxxi.41), and those who could only accept it as dogma and whose untutored passions would have to be overcome by means of carefully framed rhetorical arguments. Therefore, while examining Hobbes's writings I will make an attempt to distinguish between genuine problems in his position and inconsistencies which he consciously tolerated in order to adapt his view to the capacities of his general audience.

³ All references to the *Leviathan* are to chapters and paragraphs given in the Edwin Curley edition.

⁴ Cf. David Johnston, *The Rhetoric of 'Leviathan,'* 23.

⁵ Cf. the thesis of xlvi in light of xii.

The Causes of Crime and Criminal Responsibility

Hobbes's teaching on punishment presupposes his view of the causes of criminality, which he lays out in chapter xxvii of the *Leviathan*. There, Hobbes claims that "the source of every crime is some defect of the understanding, or some error in reasoning, or some sudden force of the passions" (xxvii.4).

By a defect of the understanding Hobbes means an ignorance of fact, such as what the law of the land requires or forbids; who enforces the law; and what are the penalties defined by law. Ignorance of this sort, Hobbes argues, can in some instances clear the offender of criminal liability. For example, if it so happens that a criminal law has not been promulgated, then no one can be held accountable for breaking it, since no one could know in such circumstances that his actions were forbidden. On the other hand, ignorance of the legally prescribed penalty cannot be an excuse for violating the law (assuming it had been sufficiently publicized) since everyone can be expected to know that punishment always follows crime, and so anyone who commits a crime without first inquiring about the legal penalty attached to it can be held responsible for his negligence (xxvii.7).

The second type of crime defined by Hobbes results from "a defect in reasoning (that is to say, from error)," by which is meant either belief in erroneous principles of right and wrong or erroneous inferences from true principles (xxvii.10-12). In earlier chapters of the *Leviathan*, Hobbes had presented a carefully constructed theory of natural right and natural law (which I will discuss in due course), but here, in the context of his

discussion of crime, he limits himself to speaking only of those principles of right that are already found in common opinion and thus known to everyone. Not everyone can be expected to know all of the specific moral laws that can be deduced by reason, but everyone can be expected to know such universally acknowledged moral rules as the „golden rule,’ “because every man that hath attained to the use of reason is supposed to know, he ought not to do to another what he would not have done to himself” (xxvii.4), and so each person can be expected to refrain from harming his innocent fellows. Likewise, everyone ought to know his duty to obey the laws of his political community (whatever the content of those laws happens to be) since each person knows that obeying the laws is more or less what we mean by being just. And yet some men ignore such commonly acknowledged rules of right and wrong and presume upon “false principles,

as when men (from having observed how, in all places and in all ages, unjust actions have been authorized by the force and victories of those who have committed them, and that . . . the weaker sort, and those that have failed in their enterprises, have been esteemed the only criminals) have thereupon taken for principles, and grounds of their reasoning, *That justice is but a vain word; that whatsoever a man can get by his own industry, and hazard, is his own; that the practice of all nations cannot be unjust; that examples of former times are good arguments of doing the like again,* and many more of that kind; which being granted, no act in itself can be a crime, but must be made so (not by the law, but) by the success of them that commit it . . . so that what *Marius* makes a crime, *Sylla* shall make meritorious, and *Caesar* (the same law standing) turn again into a

crime, to the perpetual disturbance of the peace of the commonwealth. (xxvii.10; emphases in original)

It is at first unclear how the attitude Hobbes describes in the quoted passage might be a “defect in reasoning.” Hobbes cannot mean that these men lack knowledge of the true principles of right and wrong which, after all, are known to everyone. Rather, it seems that these men know of the commonly accepted rules of morality, and yet reject them as false. The truth suppressed by conventional opinion, they think, is that all laws have their ultimate origins in crime and that justice is therefore “but a vain word.” Having liberated themselves from conventional opinion, these men believe that there is nothing further to hold them back from profiting from injustice. This is quite a sophisticated group of criminals, who espouse what one might call the conventionalist critique of justice. Contrary to Hobbes’s claim, then, they appear to be quite thoughtful, and not irrational in the ordinary sense of acting without knowledge of what they are doing. What, then, does Hobbes mean when he says that they suffer from a “defect in reasoning”?

What Hobbes seems to be suggesting here is that those who deny the truth of the common rules of morality are proved wrong by experience, and he illustrates this by his reference to the violent civil wars that plagued ancient Rome. If someone should follow the example of a Marius—whose success in committing the ultimate crime of rebelling against his government allowed him, upon acceding to power, to decree what he accomplished by violence to be just—he thereby opens the door for the Syllas and the Caesars, and others like them, to defy his own laws, to his own ruin and “to the perpetual

disturbance of the commonwealth.” Hobbes relies on the fact that these criminals are motivated by self-interest to show that their defiance of justice is imprudent and inadvisable on their very own terms. Experience shows, Hobbes contends, that while some may grow strong enough to overthrow the laws, they never live to enjoy the fruits of their labors, since by their success they bring about civil war in which the wise and resourceful are just as vulnerable to attack as the unwise and improvident, a condition Hobbes calls the “state of nature” (cf. xiii.1-3, xv.5-7). In all times, the conventionalists’ greatest error has been overlooking the fact that the superiority on which they reckon depends upon the same conventional order which they regard with contempt. The reason for their oversight is, in turn, their failure to see that in the state of nature all men return to their original equality and vulnerability. Since the actions of these conventionalists defeat their own purposes, Hobbes can say that in the strict sense they act “from a defect in reasoning,” as they are not aware of what they are doing.

Hobbes does not criticize this class of criminals for acting for the sake of their private advantage, and so in speaking about “false principles” of right and wrong in this context he really means something more like miscalculated policies of self-interest. Apparently, Hobbes thinks that the common rules of morality, which most men observe, are “true” principles because they are most advantageous. To act in accordance with reason would then mean following an effective policy of self-interest. Does Hobbes intend thereby to imply that in observing the common rules of morality most men are acting only for the sake of their private interest (e.g., to avoid punishment)? Or does he

mean to imply that when men observe these common rules out of a sense of self-sacrificial duty they are really acting out of a defect of reason?

Hobbes says that criminals belonging to the second class are still liable to punishment, despite having acted in ignorance of the consequences of injustice, ostensibly because they can be held responsible for taking pains to learn what is in their best interest (xxvii.12). But is Hobbes's claim that it is *never* advantageous to be unjust borne out by experience? It is a stretch to argue that cheating on one's taxes is likely to destabilize civil society and bring about the state of nature, and it is at least conceivable that an individual could be so politically gifted as to be able to accede to power by injustice and nevertheless maintain his rule indefinitely. Would it then be so irrational for someone with well-founded confidence in his ability to escape with impunity to commit injustice? Nevertheless, the ability to do injustice with impunity must be rare, and so a more modest version of Hobbes's position—that it is imprudent to be unjust as a general rule in light of the fragility of civil society—has a lot of merit.⁶

The third cause of crime is the passions. The one passion most likely to lead to injustice is vain-glory—the foolish overrating of one's own worth, power, or wisdom. The love of contemplating their own superiority, real or imaginary, abides in all human beings, and contributes more than anything else to their forgetting of their true interests—

⁶ Hobbes also acknowledges some important instances when individuals are released by their right to self-preservation from their obligation to obey their sovereign, but he does so without undermining his principle that it is in one's interest to be just. Since the right to self-preservation is a fundamental right and the basis of all other rights and obligations, nobody has a duty to do anything whereby he forfeits his own life, even if he should be commanded to do so by his sovereign. Thus a soldier has the right to flee danger even against his commander's orders (xv.29, xxi.16); a convicted felon has the right to escape execution (xxi.14-15); and a criminal suspect has the right not to accuse himself of a crime (xxi.13). In all of these cases, the person who disobeys his sovereign does not commit injustice since he is freed from his obligation by his fundamental natural right to preserve himself.

life and material well-being—and leads to the terrible condition of the state of nature, which is a war of all against all (cf. xiii.4-7). A man who overrates his own power or wisdom because of his excessive love of glory and does injustice because he believes himself to be strong enough to escape with impunity brings about his own ruin in the same way as the criminals belonging to the second class: he is either crushed by the stronger sovereign power or he brings about mutually destructive civil war (xxvii.13-16).

Other passions that commonly cause crime are anger, hate, lust, ambition, greed, and fear (xxvii.17-20). These, Hobbes says, are “infirmities so annexed to the nature both of man and all other living creatures, as that their effects cannot be hindered but by extraordinary use of reason, or a constant severity in punishing them” (xxvii.18). When the passion is sudden, and thus harder to control in time to prevent the unjust deed, the crime is more excusable than when the passion is long-standing. “But there is no suddenness of passion sufficient for a total excuse; for all the time between the first knowing of the law and the commission of the fact shall be taken for a time of deliberation, because [the one possessed by the passion] ought, by meditation of the law, to rectify the irregularity of his passions continually” (xxvii.33).

From his statements about criminal responsibility with respect to all three categories of crime, we can conclude that Hobbes ties criminal responsibility to private interest. Individuals who have reached maturity and taken up responsibility for their private affairs can be reasonably held accountable to the laws, since it can be assumed that their intense concern for their own welfare will extend to their taking care to observe the law, which exists for the sake of their advantage (xiv.5, xvii.1).

Moral Culpability

The three causes of crime in Hobbes's account—ignorance of facts, miscalculation of one's advantage, and the passions—are in a sense uncontroversial, since all three are acknowledged as potential causes of wrongdoing by Aristotle in his classic treatment of moral action in his *Nicomachean Ethics*.⁷ What is striking and questionable about Hobbes's account, however, is its implication that *all* injustice is ultimately due to some failure to perceive one's greatest advantage, whether due to incomplete knowledge of particular circumstances, faulty reasoning, or overwhelming passions clouding the mind. What Hobbes's account ignores is the traditional view that there is another important cause of crime: *vice*. Vicious men, we tend to believe, commit injustice not because they have miscalculated about means, but because they have a corrupt sense of what is good and choiceworthy, believing it to be something other than peaceful and lawful coexistence with their fellows.⁸ Whereas criminals acting from ignorance or lack of self-control might only merit correction, vicious criminals are thought to be deserving of a comeuppance. We tend to distinguish the two by saying that the former are criminally liable while the latter are also morally culpable. What then compels Hobbes to go against ordinary moral experience by denying the possibility of morally culpable crime?

⁷ Cf. *Nicomachean Ethics*, 1110b33-11a20 (ignorance of facts); 1111a21-35 (the passions). The possibility of miscalculation about one's advantage leading to wrongdoing seems to be implicit at 1112a18-13a9.

⁸ Cf. *ibid.*, 1113a15-b2, 1113b22-26.

Hobbes's account of criminality denying the possibility of moral culpability works by attributing all crime to some disorder in our capacity to deliberate. In all rational deliberation the end of action is presupposed. Once the end is known, we think about the most effective means that can bring it about, and then we think of the means necessary to bring about that means, and so on until in our deliberation we discover a means at the beginning of the chain that is within our power (cf. vi.57).⁹ But precisely because deliberation is about means only, it can give us no guidance regarding the ends we ought to pursue. Traditionally, it was thought that the ends for the sake of which we act both reveal what type of moral character one has—namely, how one is disposed towards the noble and the base—and determine what type of person one will be in the future. Aristotle seems to echo this outlook in Book III of his *Ethics*:

For in accordance with each sort of moral character there are special things that are beautiful and pleasant, and the person of serious moral stature is distinguished most of all, perhaps, for seeing what is truly so in each kind . . . Since then what is wished for is the end, while the things related to the end are deliberated about and chosen, the actions involving these things would be results of choice and willing acts. And the kinds of activity that belong to the virtues are concerned with these acts. Therefore virtue is up to us, and likewise also vice . . . But if doing the things that are beautiful or ugly is up to us, and likewise refraining from doing them, and this is what it is to be good or bad people, therefore being decent or base is up to us. (1113a32-14)

⁹ Cf. *ibid.*, 1112a15-b32.

. . . Still, people are themselves responsible for having become that sort [i.e., careless] by living carelessly, and for being unjust or dissipated, in the one case by acting dishonestly or, in the other, by passing their time in drinking and things of that sort, for it is the work involved in each way of acting that produces such people. (1114a5-10)

. . . For to both the good person and the bad, the end appears and is laid down in the same way, by nature or in whatever manner, and they act in whatever way they act by referring everything else to this end. (1114b14-17)¹⁰

Just as we tend to judge a person as noble or base according to the ends for the sake of which he acts, so we tend to judge the same crime to be more morally culpable when the intention behind it is base (e.g., profit) than when the intention is of a nobler sort (e.g., avenging one's honor). Hobbes tacitly denies all of this—without immediately indicating his reasons—by attributing all criminality to failure of deliberation. We must now see if there is any evidence to sustain Hobbes's denial of the possibility or significance of moral culpability.

Moral Culpability, the Relativity of the Good, and Choosing Vice

Hobbes's rejection of the notion of moral culpability receives support from his argument about the relativity of the good. In one of the most frequently quoted passages of the *Leviathan*, Hobbes denies the possibility of the greatest good, or *summum bonum*, spoken of in the books of the old moral philosophers. In place of this classical view,

¹⁰ I have made slight changes in the English translation I have consulted (Sachs's) for the sake of clarity.

Hobbes defines felicity as “a continual progress of the desire, from one object to another, the attaining of the former being still but the way to the latter” (xi.1). The reason why human life consists of a restless progress of desire is, according to Hobbes, because we ourselves are constantly undergoing change. Laying the groundwork for the psychological theories of Locke and Rousseau, Hobbes conceived of human beings as essentially malleable and as acquiring their passions, mental conceptions, and characters from the action of their physical surroundings on their organs of sense (i.1-2, 4; ii.1-2). Since the individual’s physical constitution varies in accordance with his external environment, the objects of his passions vary as well, even though the manner in which the passions operate remains roughly constant (vi.6; “Introduction,” 3). The “good,” Hobbes thought, signifies nothing more than the variable objects of our passions, while “evil” is the name we give to that for which we feel a passionate aversion (vi.7). “For these words of good, evil, and contemptible are ever used with relation to the person that useth them, there being nothing simply and absolutely so, nor any common rule of good and evil to be taken from the nature of the objects themselves, but from the person of the man . . .” (ibid.). Hobbes similarly relativizes the noble by linking it to the good. The “noble” is simply the name we give to “that which by some apparent signs promiseth good,” while “base” is the name we give to “that which promiseth evil” (vi.8).

If the ends for the sake of which we act are entirely relative to each person, a universal end, or hierarchy of ends, for human beings would be inconceivable. The implication of this for the traditional understanding of wrongdoing is profound: without a universal standard of the good it would be impossible to blame anyone for having a base

or corrupt moral character, since our moral characters are defined by our ends, which are relative. Hobbes criticizes “Aristotle and other heathen philosophers” for teaching the opposite doctrine, that private men “may judge the goodness or wickedness of their own and of other men’s actions,” when in fact “no man calleth good or evil but that which is so in his own eyes” (xlvi.32).

In the *Ethics*, Aristotle attempted to defend the position that the morally serious gentleman sees the noble and good things as they truly are by nature, and is a kind of rule and measure for other human beings whose vision of the proper ends is distorted, as is that of the vicious men.¹¹ Reflecting critically on this position shows that it may indeed be possible for the serious gentleman to possess privileged insight as to the universal standard of the good. And yet Aristotle seems unable to provide adequate proof that this is the case, leaving it a question whether the gentleman’s ends are not merely relative to his own tastes. As a result, Hobbes’s criticism of the position Aristotle attempted to defend appears comparatively strong. Not altogether undeservedly, this criticism has retained its force in contemporary liberal societies and continues to shape our understanding of moral action and wrongdoing.

Hobbes’s view rules out moral responsibility in another, quite obvious way. If each person’s ends are determined by his desires, and if the objects of his desire are shaped by the external environment, then no one can be said to choose his own ends. Aristotle had affirmed that having a virtuous or vicious character is up to us. But just as he had difficulty proving that the morally serious gentleman has privileged insight into

¹¹ Cf. 1113a15-b2.

the universal standards of what is noble and good, so he also found himself unable to give sufficient support for his claim that each person chooses his moral character.

From what I have just said, it may appear that Hobbes's denial of moral responsibility depends on his mechanistic view of human action. But while it is true that the language of behaviorist psychology is present throughout Hobbes's account of the good and the noble, the real force of his position consists rather in his implicit questioning of the conventional view that the vicious criminal willingly chooses his bad character. By defining the noble as that which is *productive* of good things (but not necessarily good *in itself*) Hobbes is expressing the common view that nobility or virtue is choiceworthy despite the fact that it may also be difficult and unpleasant, which requires no sophisticated assumptions whatsoever. By this same definition of the noble, Hobbes also forces his readers to ask themselves whether anything can truly be called noble if it is not, in one way or another, eventually conducive to our own good, even if only indirectly. This, too, seems to be in keeping with conventional opinion, for is it not true that both the poets and Scripture itself (e.g., through the story of Job) teach us that virtuous men prosper in the end? But if this conventional definition of nobility were applied with consistency, it would lead to a highly unconventional conclusion. If nobility or virtue is productive of the good, then its opposite—vice or baseness—can only be regarded as bad for us. And if vice is bad for us, what could it mean for anyone in full possession of his faculties, and with full knowledge, to choose it? Hobbes's criticism of moral responsibility can therefore be said to proceed analytically from the very way in which we speak about virtue, rather than from any questionable mechanistic premises.

The radical and debunking character of Hobbes's teaching on the causes of crime becomes still clearer in his criticism of the Christian teaching about imputing moral culpability to unjust intentions:

To be delighted in the imagination only, of being possessed of another man's goods, servants, or wife, without any intention to take them from him by force or fraud, is no breach of the law that saith *Thou shalt not covet* . . . For to be pleased in the fiction of that which would please a man if it were real is a passion so adherent to the nature both of man and every other living creature, as to make it a sin were to make sin of being a man. The consideration of this has made me think them too severe, both to themselves and others, that maintain that the first motions of the mind (though checked with the fear of God) be sins. (xxvii.1)¹²

The man who only contemplates adultery is in no way nobler in character, according to this passage, than the one who acts on his desire, since the former is only “checked with the fear of God”—i.e., by the fear of harsh punishment. But whereas Jesus, in his Sermon on the Mount, had condemned both types of men as morally culpable to an equal extent, since both have the same illicit desire, Hobbes clears both of moral culpability (though not of criminal liability), on the ground that the passions are “adherent to the nature both of man and every other living creature.”¹³ Here Hobbes is turning Jesus's Sermon on the

¹² The sentence immediately following the quoted passage is, “But I confess it is safer to err on that hand than on the other.” I have left it out because it is introduced by Hobbes without any supporting arguments and thus is likely to have been added by him to avoid the charge of heresy.

¹³ Cf. Matthew 5:21-30.

Mount on its head. Whereas, in that sermon, Jesus had presented himself as softening the Jewish Law's commandment to take an eye for an eye, by teaching that one should love his enemy and turn the other cheek, Hobbes points to the fact that the pre-Christian notion of moral culpability—as involving the agent's intention—is preserved in its entirety in Jesus's teaching. Hobbes implicitly rejects Jesus's teaching as based on a misconception about the nature of vice and wrongdoing.

Nevertheless, Hobbes did not think that his position undermined the possibility of *criminal* responsibility, as we have already seen. Having been decisively influenced by Hobbes, modern criminal law continues to recognize the importance of intention—what lawyers refer to as *mens rea*—as something that aggravates guilt when it is present, and extenuates it when it is absent. The key to understanding how Hobbes separates criminal intention from moral intention is to see that, for him, justice is not relative in the way that the good and the noble are relative to each particular person. One of the puzzles of Hobbes's political thought is trying to square his claim that all good and evil are subjective with his claim that violent death is a universal evil. But Hobbes saw no difficulty in this, believing the latter view to follow from the former. Although felicity is “a continual progress of the desire, from one object to another,” one thing will always remain constant: the preconditions of felicity will always be life itself and the power to secure oneself in its enjoyment. By default, death becomes a universal “evil” and power becomes a universal “good.” Since justice is the agreement of men to give up their natural freedom in order to better secure themselves against violent death, it acquires the same universal status in Hobbes's political theory as power and the fear of death.

Although we know that men's desires are inconstant, we *can* expect them all to be intensely concerned about their own advantage—that is, about their power and security—which means that we can also hold them responsible for being just, which, after all, is always advantageous for them.

This novel understanding of crime and criminal responsibility in turn determined the character of Hobbes's radically new theory of punishment, as we shall see in the next chapter, and of all subsequent modern penal theory until Kant. But before we can turn to the question of punishment, it is necessary to discuss an important ambiguity in Hobbes's teaching about the meaning of justice and injustice.

Justice, Self-interest, and the "Laws of Nature"

As I argued at the beginning of the chapter, Hobbes holds a unique place among modern political thinkers as the first to attempt to transform or recast the bold ideas of his modern predecessors and contemporaries into a political theory that would be morally acceptable to most, if not to all, men. Specifically, Hobbes had to show how absolute obedience to the commands of the sovereign could be not only prudent from the point of view of self-preservation, but also righteous and even dignifying. Within Hobbes's thought, this function is performed by the doctrine of the "laws of nature."

The laws of nature are precepts of reason which forbid anything that may be destructive of one's own life (xiv.3). In the *Leviathan*, there are nineteen natural laws in all, including the fundamental laws to seek peace, to constitute civil society by contract, and to perform one's covenants, as well as more specific laws against ingratitude, pride,

arrogance, and intolerance. Initially, Hobbes appears to understand these laws as deriving from the fundamental natural right of all men to resort to any means necessary for their preservation, and he expresses this by calling the laws of nature “convenient articles of peace” (xiii.14). He confirms this later by saying that the laws of nature “are not properly laws” at all, since laws properly speaking are enforceable commands of a sovereign power (xxvi.8; cf. xvii.2).

Yet Hobbes is not content to leave it at this. He moves beyond the view that the laws of nature are merely “convenient articles of peace” by claiming that although they do not bind us *in foro externo* when there is no common power to enforce them, they are *always* binding on the conscience (*in foro interno*), “that is to say, they bind to a desire they should take place” (xv.36). “And whatsoever laws bind *in foro interno* may be broken, not only by a fact contrary to the law, but also by a fact according to it, in case a man think it contrary. For though his action in this case be according to the law, yet his purpose was against the law, which, where the obligation is *in foro interno*, is a breach” (xv.37). It is one thing to think of the “natural laws” as general rules of thumb or policies which it is prudent for anyone to observe who intends his own safety. But it is entirely another thing to say that they are “obligations” always binding on the conscience, and that to disobey one of these obligations is a “breach.” We do not normally speak of “transgressing” our adopted policies when we no longer find them expedient.

We must act in conformity with the laws of nature when there is a common power to enforce them, and when there is no such power we must “desire they should take

place.” But Hobbes goes even further than this to say that we must desire to obey the natural laws *for the right reason*:

A just man, therefore, is he that taketh all the care he can that his actions may be all just; and an unjust man is he that neglecteth it . . . Therefore a righteous man does not lose that title by one or a few unjust actions that proceed from sudden passion or mistake of things or persons; nor does an unrighteous man lose his character for such actions as he does or forbears to do for fear, *because his will is not framed by the justice, but by the apparent benefit of what he is to do*. That which gives to human actions the relish of justice is a certain nobleness or gallantness of courage (rarely found) by which a man scorns to be beholden for the contentment of his life to fraud or breach of promise. This justice of the manners is that which is meant where justice is called a virtue, and injustice a vice.

(xv.10; emphasis added)

In denying that the just man acts for the sake of his own benefit or from fear of punishment, Hobbes appears to have traveled quite far from his basic teaching about the origin of justice in men’s attempt to escape the state of nature for the sake of their preservation.¹⁴ Moreover, Hobbes now implicitly adopts a hierarchy of moral characters resembling that for which he had criticized Aristotle and the heathen philosophers, since he now appears to be claiming that those who choose justice for its own sake, as their end,

¹⁴ According to his basic teaching, man’s original moral status is characterized by his absolute right or freedom—i.e., from all obligation—to do whatever may be necessary in his own judgment to protect his life, where *right*, or freedom, is strictly distinguished from *law*, or moral obligation. Cf. xiv.3.

are superior to those who act justly only for the sake of their advantage (cf. also xiv.7; xv.12, 36-37).

Whether or not Hobbes seriously held this view of justice as a moral virtue transcending self-interest, or whether he merely proposed it as a politically salutary dogma for those unable to identify justice with their own advantage, is a disputed question. Taylor and Warrender have each tried to make the case that Hobbes did take this view seriously.¹⁵ As we have already seen, however, there is also weighty evidence to the contrary. Whatever the case may be, what is important for our purpose is that Hobbes clearly meant for this noble view of justice to be taken seriously by most, if not all, of his readers, and this has important consequences for his teaching on crime. Hobbes could rule out the possibility of moral culpability only by denying the existence of universal standards of what is noble and base and by rejecting the claim that we freely choose our moral characters. But this critique of the basis of traditional morality works at cross-purposes with his insistence that the just man can and does act with a view to something other than his own advantage—namely, for the sake of justice itself—which clearly contradicts Hobbes’s hedonistic determinism. Hobbes’s critique of the possibility of universal standards of the noble is also at odds with his praise of justice as “a certain nobleness or gallantness of courage”—that is, as a moral virtue simply, rather than a virtue of merely some particular man or group of men. If societies educated by Hobbes were taught to believe that injustice is a vice and something blameworthy precisely

¹⁵ See A. E. Taylor, “The Ethical Doctrine of Hobbes”; Howard Warrender, *The Political Philosophy of Hobbes: His Theory of Obligation*.

because it is selfish, it would be only a short step from this to begin to consider unjust intentions as morally culpable, and not merely criminally liable.

For example, if a criminal on trial were proven to be animated by utter contempt for the life, dignity, and equality of human beings—things that receive the greatest sanction from the laws of nature—could a Hobbesian judge and jury be faulted for considering the accused to be morally culpable, in addition to being criminally responsible, and thus more worthy of reproach and punishment in the eyes of the law? Hobbes even seems to encourage such a view in his discussion of crime by asserting that the gravity of a crime ought to be measured by “the malignity of the source or cause,” as well as by its harmful consequences (xxvii.29). In retrospect, we can also see that the ambiguous way in which Hobbes refers to the second cause of crime—error regarding “principles” (rather than “policy”)—allows for a more traditionalist interpretation of his teaching on criminality. With this difficulty in mind, I will now turn to examine Hobbes’s theory of punishment.

Chapter 2: Hobbes's Moral Theory of Punishment

Retribution in the Modern Context

The most significant innovation of Hobbes's theory of punishment is its wholesale rejection of retribution. Retribution is an essentially backward-looking response to crime inspired by a feeling of righteous indignation towards those responsible. Specifically, the aim of retribution is to subject the offender to suffering in proportion to the viciousness of his character, as it is revealed in the seriousness of his crime and the intention with which he committed it—this is what we tend to call a punishment's "fittingness." Consider the reaction of most people to a tyrant when he is brought down: they call for his head. This reaction is not primarily motivated by a desire to deter other would-be tyrants, or to put the tyrant out of his misery as if he were suffering from a sickness, but by a concern with retributive justice. Most people do not believe that a tyrant acts out of a pathological condition he does not control (as does a serial murderer who kills in order to cannibalize his victims) but out of a vicious lack of moral scruples. It is true that I have not provided a thorough account of the meaning of vice or how it differs from mental illness, and so it might seem as though I am presuming too much by distinguishing vice from pathology. But I do not believe that such an account is necessary at this juncture (although it will become necessary in later chapters). On the contrary, I believe it is much less presumptuous, and much more useful in talking about wrongdoing in ordinary practice, to draw such a distinction.

We are prone today to an unfortunate prejudice—induced by Rousseau but not shared by him—that acts of injustice, and especially the greatest crimes, are due to an undeveloped sense of compassion on the part of the criminal, as a result of his having been denatured early in his life by an unfortunate upbringing. I used the example of the tyrant because it makes much more sense to attribute his actions to vice (in the ordinary, unsophisticated sense of the word) than to an absence of compassion. Does the tyrant dominate others because he simply cannot imagine their suffering, and thus does not realize what he is doing? This seems highly improbable.¹ It is much more probable that the tyrant suppresses his feeling of compassion through an unusually strong will-power, so that he can pursue his own advantage at the expense of everyone else. And is this calculating disregard of justice not what we ordinarily refer to as vice?

Classical political philosophy had been skeptical of retributive punishment, and yet refrained from openly criticizing it in principle and generally sought to accommodate it in practice. We can observe how this occurs in Aristotle's *Nicomachean Ethics*. In his discussion of the virtue of justice in Book V, he names distributive justice and corrective justice as the only specific types of justice, never mentioning retribution. Aristotle defines corrective justice as concerned with rectifying (according to arithmetic proportion) any imbalance in a prior distribution of goods, where the role of the judge is to take back an unjust gain and to restore an unjust loss. This includes all types of unjust exchange among citizens, including crime, which Aristotle defines as a specific type of exchange ("unwilling" exchange) since it often involves a gain for the offender and a loss for his

¹ As Rousseau says himself. See "Discourse on the Origin and Foundations of Inequality among Men," in *The First and Second Discourses*, 131.

victim.² But Aristotle also admits that it is wrong to apply the economic terms “gain” and “loss” to certain types of crime, since, for example, the man who committed an assault out of anger or hatred has not really gained anything of benefit to himself.³ Aristotle implies that adding to the assaulter’s punishment anything beyond what is required for compensation of the victim’s loss (and possibly deterrence) because we consider the assaulter as having gained something from his assault would not be just. But later he admits that a political community stays together “by paying things back proportionately, since people seek either to pay back evil, and if they cannot that seems to be slavery, or to pay back good,” and in speaking of paying back evil, he seems to be referring to retributive punishment.⁴ Thus Aristotle seems to make a reluctant concession to the ordinary civic outlook, which demands retributive justice in addition to corrective and distributive justice—perhaps because he realized that it receives strong support from his own endorsement of moral responsibility earlier in the *Ethics*.

Unlike Aristotle’s cautious and accommodating treatment of retribution, Hobbes’s emphatic rejection of moral responsibility and his assertion of the relativity of the good make retributive punishment simply inconceivable. To use again my example of the man with tyrannical ambitions, it is impossible on a Hobbesian view to condemn such a man for being morally wicked; one can only admonish him for having set out on a dangerous course that is likely to fail and to try to force him to desist. Consequently, the seventh law of nature is:

² *Nicomachean Ethics* 1131b25-32a8.

³ *Ibid.*, 1132a6-14

⁴ *Ibid.*, 1132b31-33a1; cf. 1113b23-25.

That in revenges (that is, retribution of evil for evil) men look not at the greatness of the evil past, but the greatness of the good to follow. Whereby we are forbidden to inflict punishment with any other design than for correction of the offender, or direction of others. For this law is consequent to the next before it [i.e., the sixth law], that commandeth pardon upon security of the future time. Besides, revenge without respect to the example and profit to come is a triumph, or glorying, in the hurt of another, tending to no end (for the end is always somewhat to come); and glorying to no end is vain-glory, and contrary to reason; and to hurt without reason tendeth to the introduction of war, which is against the law of nature, and is commonly styled by the name of *cruelty*.
(xv.19; emphases in original)

Here in a nutshell we have the principle of all subsequent modern theories of punishment prior to Kant. Although in this passage Hobbes calls the kind of punishment he advocates “retribution” or “revenge,” he defines it as looking only to the future, and thus destroys the original meaning of retribution. According to Hobbes’s view, the only legitimate uses of punishment are those which produce some future benefit. This can include correction of the offender, if this is possible, or incapacitation, if it is not, as well as punishing the convicted criminal as an example to deter others. All other uses of punishment are against reason and the law of nature, and can only be considered acts of vain-glory or cruelty.

Up to this point in my whole discussion of Hobbes, I have only noted the negative implications of his view for the meaning of crime and punishment. By eliminating the essential conditions of moral culpability and retribution—responsibility for one’s

character and the possibility of moral judgment in light of a universal standard—Hobbes renders these concepts senseless. But what now emerges from the passage I just quoted is the positive or edifying side of Hobbes’s theory of punishment. Retribution is not only rejected as incoherent, it is also forbidden as *cruel* and *vain*. In this way, Hobbes’s natural law against retribution is not simply a logical refutation, but a humane and benevolent principle denouncing cruelty and barbarism. In the following two sections, I will examine this humane side of Hobbes’s rejection of retribution in the context of his theory of punishment.

The Reasonableness and Safety of Deterrence

Hobbes begins his chapter “Of Punishments and Rewards” by defining punishment as “an evil inflicted by public authority on him that hath done or omitted that which is judged by the same authority to be a transgression of the law, to the end that the will of men may thereby the better be disposed to obedience” (xxviii.1). This definition straightforwardly reflects the seventh natural law against backwards-looking revenges by restricting the scope of punishment to deterrence. For the bulk of the remainder of the chapter, Hobbes draws out the specific consequences of this definition for the proper administration of punishment (as well as the dispensing of rewards) by the public authority. Throughout, Hobbes’s aim is to show that a strict, but also fair and humane, criminal justice system is a much more expedient policy for the enlightened sovereign than is a Machiavellian policy of bloody pacification.

Before Hobbes proceeds to the main task of his chapter, he takes care to clarify the basis of the right to punish:

Before I infer anything from this definition, there is a question to be answered of much importance, which is: by what door the right or authority of punishing in any case came in? For by that which has been said before, no man is supposed bound by covenant not to resist violence; and consequently, it cannot be intended that he gave any right to another to lay violent hands upon his person. In the making of a commonwealth, every man giveth away the right of defending another, but not of defending himself. Also, he obligeth himself to assist him that hath the sovereignty in the punishing of another, but of himself not . . . It is manifest therefore that the right which the commonwealth (that is, he or they that represent it) hath to punish is not grounded on any concession or gift of the subjects.

But I have also showed formerly that before the institution of commonwealth, every man had a right to everything, and to do whatsoever he thought necessary to his own preservation, subduing, hurting, or killing any man in order thereunto. And this is the foundation of that right of punishing which is exercised in every commonwealth. For the subjects did not give the sovereign that right, but only (in laying down theirs) strengthened him to use his own as he should think fit, for the preservation of them all; so that it was not given, but left to him, and to him only, and (excepting the limits set him by natural law) as entire as in the condition of mere nature, and of war of every one against his neighbor. (xxviii.2)

The argument put forward here needs no clarification. Hobbes merely draws out the necessary implications of the basic elements of his political theory: since the natural right to defend one's life and limb is prior to, and the foundation of, all obligations, it is inconceivable that anyone could grant another the right to do him physical harm for any reason. Consequently, the right to punish must be grounded in the absolute natural freedom which individuals in the state of nature retain until they renounce it in order to constitute civil society, and which the sovereign always retains unimpaired since he is not a party to the social contract (cf. xvii.13).

But in the very same breath, Hobbes observes that the sovereign's right to punish is left whole *excepting the limits set him by natural law*. This paradoxical position, that the right to punish is both unlimited and limited, reflects Hobbes's attempt to reconcile the two divergent trajectories which his thought follows. On the one hand, Hobbes never loses sight of the fundamental purpose for which modern government is established—i.e., to forestall or quell all political and moral dispute—and this requires that sovereignty should remain essentially unrestricted. Hobbes expresses this commitment to absolute sovereignty by grounding the sovereign's right to punish in the unlimited right of nature. At the same time, Hobbes wishes to reassure subjects that the exercise of sovereign power, particularly in the form of law-enforcement, will tend to be guided not by the passion or whim of the sovereign but by a higher moral law. Hobbes is walking a fine line here: he accepts that the sovereign is absolutely unrestricted by any other *power* while still being *morally* bound by the law of nature.

At first, this seems a small consolation to subjects who would have no recourse should their sovereign choose simply to ignore his solemn obligations under the laws of nature. And yet Hobbes's distinction is not mere window-dressing for the sovereign's arbitrary exploitation of his subjects. Even if he is not guided by the best of motives, the sovereign will still be inclined to obey the natural law for the simple but powerful reason that it is in his own best interest to do so. This is illustrated by Hobbes's specific inferences from his definition of punishment. The fifth inference forbids all punishment "which is inflicted without intention or possibility of disposing the delinquent (or by his example, other men) to obey the laws." According to Hobbes, this is not strictly speaking punishment at all, "but an act of hostility" (xxviii.7). Initially, this inference seems to be a limitation on the sovereign's freedom to act, but it is in fact no such thing: by punishing for the sake of vengeance, and so inflicting harm "without intention or possibility of disposing the delinquent . . . to obey the laws," the sovereign does himself no service. Assuming that the sovereign's most abiding concern is to rule securely over an obedient population, he would be well advised to refrain from punishing his subjects for any reason other than deterrence, as a direct corollary of Hobbes's definition of deterrence. Moreover, insofar as the sovereign does limit his aim to deterrence, it is good policy to impose no more than the minimum penalty required for effective deterrence, as more than this would be superfluous; and no less than the minimum, since that would not be a punishment but the "price" of committing the crime, serving not to dispose men to obey the law but encouraging them to break it (seventh inference; xxviii.9).

The eighth inference from Hobbes's definition of punishment states that a penalty should not exceed the amount of harm stipulated *by law*. "For seeing the aim of punishment is not a revenge, but terror, and the terror of a great punishment unknown is taken away by the declaration of a less, the unexpected addition is no part of the punishment" (xxviii.10). In speaking of extralegal punishments, Hobbes likely has in mind punishments inspired by the desire for revenge, since impassioned punishments are the most likely to be impulsive and unpredictable. Punishments that are not announced in advance by law are not advantageous for the sovereign because they are unable to deter. Moreover, by their very unpredictability they may actually undermine obedience by making punishment seem more willful and giving would-be criminals hope that it might be avoided by appealing to the sovereign's (or the judge's) favor (cf. xxvii.32). Hobbes was therefore clearly aware of the implication of his theory—which has become a truism for contemporary criminologists—that the effectiveness of punishment depends on its certainty just as much as on its severity (cf. xxvii.18).

The third inference states that offenders have a right to public trial before they can be punished (xxviii.5). Any punishment that is not preceded by a public trial has no ability to deter, since no one besides the sovereign and his ministers can know what law the punished criminal had broken. For the same reason, the ninth inference from the law of nature forbids inflicting punishment for an action that was criminalized retrospectively—i.e., after the fact (xxviii.11). Where no law exists prohibiting a given action, punishment cannot be anticipated and thus has no power to deter. Once again, the law of nature granting immunity from *ex post facto* laws only reflects what is already

expedient for the sovereign. Along the same lines, Hobbes also denies that pre-trial incarceration can be considered punishment (xxviii.20), and denies the right to punish the innocent (xxviii.22).⁵ The sovereign's prudential reason for observing these restrictions in administering punishment is, therefore, a great guarantee of security to law-abiding subjects, both from would-be criminals and from the sovereign himself.⁶

The Morality of Deterrence

In addition to their usefulness, Hobbes's principles for the administration of penal justice also take on the character of natural law, from which they derive. This second aspect is as important for understanding Hobbes's contribution to modern punishment theory as the first, and perhaps even more so given the subsequent development of penal thought. It bears repeating that Hobbes's natural laws are not only "convenient articles of peace," but also moral duties absolutely binding the conscience. As sharing in the majesty of natural law, Hobbes's principles of punishment take on an entirely different dimension.

Hobbes's study of the human passions taught him that men care about their dignity just as much as, and sometimes more than, about their material interests. Hobbes saw that in the important matter of the sovereign's use of coercion against his subjects, it

⁵ This, of course, does not preclude the possibility that intentionally harming the innocent in certain circumstances may help to deter crime (for example, as part of a Machiavellian policy of domination). Hobbes's official position is that punishing the innocent is morally wrong, since it is against the law of nature. But one wonders whether punishing the innocent would still be forbidden, on Hobbes's view, in a case in which this becomes necessary in order to stabilize an extremely volatile political situation, in accordance with the fundamental law of nature to seek peace.

⁶ Although not directly related to punishment, Hobbes had argued—also on prudential grounds—for the right against self-incrimination, and is therefore the father of the U.S. Constitution's Fifth Amendment protection. Cf. xiv.30.

would not be enough to show that a deterrent penal system would be the safest; it would also have to be shown that such an arrangement is compatible with the dignity of subjects—of criminals just as much as of non-criminals. The understandable tendency to focus on Hobbes’s individualism has sometimes caused readers to overlook his brief but prescient indications of the new grounds upon which civic friendship and a sense of community can flourish in a thoroughly modern society. I am speaking of men’s natural equality and similarity (cf. xv.17, 21-22, 35). Contrary to some uncharitable readings of Hobbes,⁷ the kind of society he imagined would indeed be inspired by a sense of community or fellow-feeling, though certainly not one as intense as that of a traditional community. Because of this expectation, Hobbes could write regarding the formation of civil society that “it is more than consent, or concord; it is a real unity of them all” (xvii.13). Men who depend on each other for their preservation and well-being, and who acknowledge their natural equality and common need for security, will be far from indifferent to each other’s suffering and humiliation. Criminals, as distinguished from enemies, remain members of civil society despite having offended against its laws, and are still entitled to respect as fellow citizens and co-signatories to the social contract. As Hobbes’s new account of crime rules out immoral character as a cause of criminality, there would be even less reason to treat criminals with contempt in modern societies than there had been in traditional societies. It was essential, then, for Hobbes to show that punishment for the sake of deterrence could be not only effective but also humane.

⁷ See, for example, C. B. Macpherson, *The Political Theory of Possessive Individualism: Hobbes to Locke*.

Many of the inferences in Hobbes's chapter on punishment reflect this concern. In discussing the third, fifth, seventh, eighth, and ninth inferences, I observed how their apparent restrictions on the sovereign's right to punish were not really restrictions because they only commanded the sovereign to do what was already expedient for him. But as natural laws, these inferences also forbid him to punish whenever there is no chance to deter because it would be *cruel* and *vain*, and not just because it would be imprudent (cf. again xv.19). According to Hobbes, subjects authorize all of their sovereign's actions, including the punishment of criminals. In a community which punished its criminals only with a view to deterrence, and never for revenge, a subject could feel proud of the fact that the penal system administered by his sovereign representative, and authorized by himself, was humane.

Besides once and for all repudiating cruel and vengeful punishments, Hobbes also contributed to the humanizing of modern punishment by creating the conditions for a shift of focus to rehabilitative punishment—although not without dramatically transforming its meaning. In elaborating the rules to which a rational penal system should adhere, Hobbes explicitly focused on deterrence. Yet his definition of deterrence—"that the will of men may thereby be better disposed to obedience"—is capacious enough to include rehabilitation as one of its forms, as long as its ultimate object remains crime prevention. Hobbes was more inclined to place the responsibility of keeping one's passions in check on the individual himself, and so he did not feel compelled to stress the state's potential role in dispensing non-punitive preventive care to past offenders who are predisposed to crime by psychological illness, drug addiction, or dysfunctional family

settings, although Hobbes's principles are entirely favorable to reforming criminals through prison labor, drug rehabilitation programs, or community reintegration programs—as long as this contributes to, and does not undermine, the deterrent effect of punishment.⁸ Thus it was Hobbes who laid the groundwork for the penitentiary movements of the late 18th and 19th centuries in Europe and the United States.

These humane features of Hobbes's punishment theory continue undiminished to inform modern penal thought and practice. As an illustration of this, one might consider the passionate appeals to the decency and humanity of the American people by recent penal reformers like psychologist Karl Menninger and former U.S. Attorney General Ramsey Clark, whose moralizing eloquence appears side by side with their calls for Americans to consider their own safety and security.⁹

But unlike many of the penal reformers in the Hobbesian mold (particularly the utilitarians), Hobbes was aware of one very serious, and unavoidable, moral defect or blemish of modern punishment, as such. The end for which civil society is constituted is peace and common defense—or more accurately, the defense of *each* of its members. This implies that it would be contrary to the purpose of the social contract for anyone to renounce his right to defend himself or to transfer that right to his sovereign (xiv.29). The right to defend oneself is the first and most important inalienable right. Accordingly, Hobbes acknowledges that a man convicted of a capital crime has the right to resist the sovereign's attempt to punish him, and that he does no injustice by his resistance,

⁸ According to Hobbes, the vast majority of men are already either mad or teetering on the brink of madness. See viii.15-23.

⁹ See Ramsey Clark, *Crime in America*; Karl Menninger, *The Crime of Punishment*.

notwithstanding that all of the sovereign's commands are just (xxi.12, 15). It would seem to follow from this logic that the social contract itself becomes void if or when it ceases to serve the purpose for which it was formed, as is necessarily the case for the offender convicted of a capital crime. It would follow that as soon as the sovereign makes known his intention to execute him, the convict immediately reclaims all of his original natural rights and so returns to the state of nature. This would be a significant setback for Hobbes's efforts to show that punishment grounded in modern principles could be more moral than traditional forms of punishment, since it would seem to follow from his view that capital punishment is indistinguishable from an act of war, as both bring about a return to the state of nature.

What makes matters worse for Hobbes is that he conceived of the natural right to defend ourselves to be quite expansive, including not only the right to resist others' overt attempts to kill us, but to resist any attempt to restrict our power or liberty in such a way as to make us incapable of defending ourselves against potential attempts on our life in the future. Thus, according to Hobbes, each has the right to resist his sovereign's command to punish him in all cases involving death, wounding, and incapacitation (including incarceration):

As, first, a man cannot lay down the right of resisting them that assault him by force, to take away his life, . . . the same may be said of wounds, and chains, and imprisonment, both because there is no benefit consequent to such patience . . . as also because a man cannot tell, when he seeth men proceed against him by violence, whether they intend his

death or not. [Third] and lastly, the motive and end for which this renouncing and transferring of right is introduced, is nothing else but the security of a man's person, in his life and in the means of so preserving life as not to be weary of it. (xiv.8)¹⁰

With this expansive definition of the right to self-preservation Hobbes directly anticipates Locke. Like Locke, Hobbes indicates that in a significant number of possible cases of government coercion the threatened individual finds himself suddenly transported back to the state of nature. Even if in such cases coercion should happen to be justified by criminal guilt and prior conviction, the unsettling consequence of the modern idea of inalienable natural rights is that punishment creates a very real disruption in the civil order, a kind of temporary pocket of lawlessness where nothing can be just or unjust from the point of view of the criminal.¹¹

Yet Hobbes refused to draw the conclusion that corporal and carceral punishment precipitate a return to the state of war between the criminal and civil society, even though this appears to follow directly from his basic theoretical outlook. Instead, Hobbes tries to maintain that the criminal authorizes *his own* punishment and thus remains to the end a consenting member of civil society. The first and fourth inferences of the natural law

¹⁰ In the earlier *Elements of Law*, Hobbes had maintained that each subject *does* renounce his right to defend himself against his sovereign, and thus avoided the conflict between natural right and the right to punish. See xix.10, xx.7. But that earlier position was even more damaging to Hobbes's overall purpose, since it compromised the privileged moral status of self-preservation, and so he was forced to revise his view in the later *De Cive* and *Leviathan*.

¹¹ Norrie goes too far, however, when he concludes that this "leads to the immanent [*sic*] collapse of, and implicit denial of the possibility of, the social state and the institution of punishment." "Hobbes and the Philosophy of Punishment," 307. In my view, Gauthier and Martinich are more correct to argue that the clash of the rights of the criminal and of civil society, respectively, does not create an inconsistency in Hobbes's political theory, although it does reveal the limits of the social contract. See David Gauthier, *The Logic of Leviathan*, 146-49; Aloysius Martinich, *Hobbes*, 115-18.

regarding punishment state that any evil inflicted by anyone other than the sovereign is an act of hostility, “because the acts of power usurped *have not for author the person condemned,*” as they ought to have if they are to be considered punishment in the strict sense (xxviii.3, 6; cf. xviii.3). But as Hobbes’s perspicuous argument about the origins of the right to punish showed, the sovereign’s right to punish does *not* rest on the consent of the punished but on his own unimpaired natural right to all things. If the right to punish is not acquired but natural, then the sovereign is the only author of his exercise of this right. How, then, could the criminal *also* be said to authorize his own punishment?

It might be argued that Hobbes was forced to conclude that the criminal is punished by his own authority as a consequence of his view that each subject authorizes all of his sovereign’s actions without exception by virtue of the original social contract (xvii.13, xviii.6, xxi.10, 14). This naturally leads us to ask why Hobbes insisted that the subject must authorize all of the sovereign’s actions in the first place, only to discover that the reason for this, too, is mysterious. If the end of civil society, and thus of sovereignty, is to defend each individual, then shouldn’t subjects authorize their sovereign’s actions only to the point that they are intended to serve this purpose? But surely not all of a sovereign’s actions are intended to serve this purpose, as, for example, his choice of what to eat for dinner or which way to part his hair; it would be absurd to say that such actions are authorized by his subjects. At any rate, corporal and carceral punishments do not serve the end of self-preservation from the point of view of the criminal, since, as Hobbes argued, any attempt to weaken us or put us in bonds diminishes our ability to defend ourselves and thus indirectly puts our lives at risk. Why,

then, does Hobbes insist that the criminal is author of his own punishment if this is not a conclusion made necessary by his own principles?¹²

At one point, Hobbes tries to make sense of his awkward position according to which the criminal is punished by his own consent *and* still retains his natural right to preserve himself:

Again, the consent of a subject to sovereign power is contained in these words *I authorize, or take upon me, all his actions*, in which there is no restriction at all of his own former natural liberty; for by allowing him to kill me, I am not bound to kill myself when he commands me. It is one thing to say *kill me, or my fellow, if you please*, another thing to say *I will kill myself, or my fellow*. (xxi.14; original emphases)

This argument turns on a specious distinction between (1) a subject's authorizing the sovereign to kill him along with the means to do it, and (2) authorizing his death without the means to accomplish it. Hobbes claims that the second may be authorized but not the

¹² Johnston argues that Hobbes's theory of authorization is a rhetorical strategy, observing that the theory of authorization only emerges in the *Leviathan*—the most rhetorical of Hobbes's three major political works—and that it is absent in the earlier *Elements of Law* and *De Cive*, suggesting that Hobbes believed it to be unnecessary. See *The Rhetoric of 'Leviathan,'* 80-82.

Cattaneo and Norrie both believe that the subject's unconditional authorization of his sovereign's actions *is* necessitated by Hobbes's principles, but I do not think that either gives sufficient reasons for this. Both rely too much on Hobbes's assertions, but do not consider whether those assertions are substantiated on Hobbes's own terms. See Cattaneo, "Hobbes's Theory of Punishment," 293; Norrie, "Hobbes and the Philosophy of Punishment," 304-6. Tuck gives a somewhat stronger defense of the necessity of unconditional authorization. According to Tuck, the sovereign's natural right only permits him to act in *his own* defense, and this appears not to give him the right to punish in order to defend *civil society* as a whole; unconditional authorization therefore becomes necessary to justify punishment. Richard Tuck, *Natural Rights Theories*, 129-130. But the problem is not as great as Tuck believes. As I have tried to show, the sovereign's personal safety and the security of civil society are mutually dependent, thus punishing the guilty is *also* an act of self-preservation by the sovereign. I have already indicated in the text the other reasons why I think unconditional authorization does not follow necessarily from Hobbes's contractarianism.

first. It is a specious distinction because it is hard to imagine how one can truly consent to the end without also consenting to the means to bring about that end. According to Hobbes, a man always does everything in his power to prevent harm from coming to himself—whether it be harm that he might cause himself or whether the harm should come from another—and so he never truly wills the means to harm himself. And if he cannot will the means, then he can never truly consent to the end either, or if he did it would be a meaningless and merely verbal consent. It is impossible to know whether Hobbes believed this distinction to be sound, or whether he knew it to be specious and was trying to win more support for absolute sovereignty by making all of its actions appear to be entirely consensual from the point of view of every subject, even the man on death row. Could it be that Hobbes wished to avoid the chilling effect of admitting that civil society must occasionally wage war against some of its members, even at the cost of logical consistency? Whatever the case may be, Hobbes makes punishment appear to take place entirely within the civil order, rather than outside of its boundaries, contrary to his own principles. As we shall see in the next section, this particular inconsistency in Hobbes's position was successfully attacked by the liberal penal thinker Beccaria in the 18th century.

We now have before us Hobbes's theory of punishment, which was also the first theory of punishment constructed entirely on modern principles. What makes it remarkable, and what sets the tone for all subsequent modern penal thought, is its attempt to show that a penal system aiming at deterrence is not only highly effective but also humane, where the latter is not simply reducible to the former. Both aspects are essential

to it, and so it would be a mistake to classify Hobbes as a simple „consequentialist.’ But how happy is this marriage between utility and humanity? How persuasive is Hobbes when he assures us that the moral treatment of criminals never requires anything more than what is useful? This question fundamentally informed the subsequent development of modern penal theory.

The Problem of the Death Penalty

For the most part, Hobbes’s basic principle that punishment ought to aim at deterrence (broadly understood) has remained unchallenged by subsequent modern thinkers (albeit with a few important exceptions). Following Hobbes, Locke declared that the only legitimate purpose of punishment is “reparation and restraint.”¹³ When Montesquieu spoke of “the just proportion between the penalties and the crime,” he said nothing of fitting the severity of the punishment to the moral gravity of the offense, but only that “it is essential that the greater crime be avoided rather than the lesser one, the one that attacks society rather than the one that runs less counter to it.”¹⁴ Less ambiguously, John Stuart Mill stated that “the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others.”¹⁵ This principle remains highly influential today under the name of „consequentialism,’ although it is now commonly divested of its connection to natural

¹³ *Second Treatise of Government*, § 8.

¹⁴ *Spirit of the Laws* bk. 6, ch. 16. See also *ibid.*, bk. 6, ch. 9 and bk. 10, chs. 2-3.

¹⁵ *On Liberty*, 9.

law which, in Hobbes's view, is what gave it its dignity. In the second half of the 18th century, however, certain features of Hobbes's theory of punishment came under attack.

The first significant challenge to Hobbes's teaching on punishment was advanced by the Italian Cesare Beccaria, one of the earliest writers to espouse the principle of utilitarianism—"the greatest happiness of the greatest number." In 1764 Beccaria published his highly influential treatise, *Of Crimes and Punishments*, which is widely credited with inspiring the so-called classical school of criminology. In this essay, Beccaria accepts Hobbes's view that the scope of punishment ought to be limited to deterrence. But Beccaria also clearly saw what was entailed by Hobbes's contractarianism: in renouncing part of their natural liberty to form civil society, men cannot possibly renounce their liberty to defend themselves, and so they cannot give their sovereign the right to kill them.¹⁶ "The penalty of death is not therefore a *right* . . . but a war of the nation against a citizen; because it has been judged necessary or useful to destroy him."¹⁷ On this basis, Beccaria called for the abolition of capital punishment as unjust.

As we saw, Hobbes had avoided drawing the conclusion that capital punishment is indistinguishable from an act of war, although this had been implied by his own principles. Beccaria's criticism of Hobbes to this extent is therefore correct. But Beccaria was not necessarily correct when he concluded that capital punishment is "not . . . a right." Hobbes had shown that the social contract binds *subjects* to one another and to their sovereign; it does not bind the sovereign himself, who retains intact his natural right

¹⁶ Cf. Leo Strauss, *Natural Right and History*, 197.

¹⁷ *Of Crimes and Punishments*, ch. 16.

to all things, and thus also the right to kill for the preservation of civil society.¹⁸ Beccaria relied on a view of the social contract the elements of which were essentially Hobbesian. He accepted that the public authority is constituted by private individuals giving up some of their natural liberties to a single representative in order to protect the remainder of their natural liberties, which they retain even while in civil society. He saw just as well as Hobbes did that so long as one allowed individuals in civil society to retain their basic natural rights—i.e., to self-preservation and to the basic means of a comfortable existence—any imposition of corporal or carceral punishment would be considered as conflicting with those rights. But Beccaria wished neither to accept this conclusion, as Hobbes did (albeit with reluctance and equivocation), nor to move in Rousseau’s totalitarian direction of demanding the complete alienation of all natural rights as the condition of civil society.¹⁹ Beccaria was therefore left in a quandary, the solution to which he believed to have found in the principle of the greatest happiness of the greatest number.

Beccaria mistakenly believed that the principle of the greatest happiness of the greatest number could be derived from the social contract:

If every individual member is bound to society, society is equally bound to every individual member by a contract binding by its nature on both parties. This obligation,

¹⁸ Hobbes understood correctly that the purpose of the social contract is to protect the contracting parties *from each other* by weakening all but one individual (or assembly), who must therefore be exempted from the social contract. It would be contrary to this purpose to then weaken the sovereign by making his authority conditional, since in his weakened state it would be more difficult for him to defend his subjects from one another. Cf. xvii.13 and xviii, entire.

¹⁹ Cf. *On the Social Contract*, Part I, chs. vi-vii.

which descends from the throne to the hovel and binds equally the greatest and the most miserable of men, means exactly this: that it is in *all* men's interest for agreements beneficial to the *greatest number* to be observed.²⁰

Beccaria thought that all deterrent punishments short of death could be justified in this way. If the principle of general utility is implied by the social contract, then since all men in civil society consent to the social contract they would have no choice but to acknowledge the justice of any action beneficial to the majority, even if their own private interests should thereby suffer. Life imprisonment, for example, could be successfully defended as providing general security for the majority, while only harming a small minority (the life-term prisoners). In this way, the principle of general utility seems to accomplish what Hobbes had failed to accomplish with his fiction of unconditional authorization—i.e., it seems to prove that any deterrent punishment (short of death) is justified even in the eyes of the criminal himself.²¹ It is not hard to see that Beccaria's argument rests on a non-sequitur. Contrary to his claim, it is *not* in the interest of *all* to observe agreements beneficial to the *greatest number*, and so the principle of the greatest happiness of the greatest number does not follow from the social contract. At most, the principle of general utility can be a useful government policy, but it cannot impose a moral obligation capable of binding a criminal to submit to a punishment that would put

²⁰ *Of Crimes and Punishments*, ch. 4, n. 1; emphases added. Cf. ch. 2, n. 1.

²¹ John Stuart Mill and Jeremy Bentham, the two more famous utilitarians, followed Beccaria in accepting the principle of general utility as sufficient to justify punishment.

his ability to exercise his natural right in jeopardy, as long as we agree that the individual's natural right is prior to, and the basis of, all possible obligation.

Incidentally, if Beccaria's principle were granted it would not be sufficient to rule out capital punishment. Beccaria attempted to argue that the institution of capital punishment can *never* be more beneficial to the greatest number than other types of punishment, such as long-term incarceration. In this way, he hoped to avoid in practice a conflict between the principle of general utility and the individual's natural right to self-preservation. Of course, such an eventuality cannot be ruled out in principle, and so Beccaria fails to prove that capital punishment is unjust.²²

Unfortunately, Beccaria's essay became instantly popular in Europe, England, the United States, and even in the court of the Empress of Russia, Catherine II. The popular victory of Beccaria's principle of the greatest happiness of the greatest number cast into obscurity Hobbes's argument for locating the basis of the right to punish in natural right. This victory thereby obscured the connection between the true philosophical basis of modern politics and actual modern penal practice, which, nonetheless, continues to be overwhelmingly concerned with deterrence. Liberals who are serious about individual rights are understandably skeptical of utilitarianism in general, and uncomfortable with its implications for punishment in particular. It is no surprise that societies animated by a shared respect for individuals find it difficult to accept that the rights of the individual

²² Bentham follows Beccaria in arguing that the death penalty is less useful as a general deterrent than other kinds of punishment, and thus should be abolished. See Bentham, *The Rationale of Punishment*, Part II, ch. xii. J. S. Mill, on the other hand, argued the reverse, that the death penalty is (paradoxically) less harmful to the criminal than life incarceration, and thus more justifiable on the principle of general utility. See Mill's 1868 speech on capital punishment, excerpted in *Utilitarianism*, 65-71.

must be sacrificed to the interest of the majority. Nevertheless, Beccaria's successful transformation of Western penal thought helpfully points to an important problem inherent in Hobbes's original conception of punishment.

Hobbes had intended to rescue modern sovereignty from infamy, and part of his project was to show that penal systems constructed entirely according to modern principles could be effective as well as moral. Hobbes tried to show that the morality of punishment could never be at odds with its effective administration. That is, Hobbes denied that the moral treatment of criminals could ever require even a small sacrifice of civil society's security. We can now see that although Hobbes had been very consistent in tracing the implications of the principles of modern politics for punishment, he failed to persuade the vast majority of his readers that the sacrifice of individual criminals, however dangerous, for the mere safety of civil society could be in complete harmony with justice. We can add that Hobbes failed in his self-imposed task in yet another way: he failed to persuade many of his readers that retributive punishment is immoral.

The Problem of Retribution

Hobbes had taught that the goods for the sake of which human beings act are relative to each agent's physical constitution. On this view, there is no highest universal good, or *summum bonum*, in light of which we might be judged as to our moral worth. These are precisely the conditions presupposed by retribution, which looks backward to the moral quality of our actions rather than forward to their consequences, and so Hobbes believed he could reject retribution on his terms. As he could find no moral foundation

for retribution, Hobbes concluded that the demand for retributive punishment that existed in conventional morality was rooted in the irrational passion of vain-glory. Modern punishment, as Hobbes conceived of it, could be regarded as morally superior to the conventional principle of punitive justice because it limited the scope of punishment to deterrence and excluded revenge as unnecessary and cruel.

Yet Hobbes refused to extend his relativist critique of the *summum bonum* to justice. The meaning of justice is not subjective but universal, as everyone, according to Hobbes, acknowledges. Making covenants and keeping them is the only means by which human beings can secure their lives and livelihood, and this remains true despite the diversity and mutability of the objects of men's desires. Moreover, as we have seen, Hobbes argued that being just is not only generally advantageous, but that it is also a moral imperative binding all men unconditionally in their consciences. A just man is not only prudent, according to Hobbes, but *noble*, as his obedience to the law is determined by his virtuous character rather than by his fear of punishment or hope for reward. In this way, Hobbes brought his theory of moral agency back to the point where it began to resemble the traditional view—as it had been articulated by Aristotle, for example. If by Hobbes's own admission justice is a universal standard and a moral virtue, and if we can judge a man's character based on the way he is disposed towards justice, then perhaps Hobbes was wrong when he concluded that retributive punishment could not be justified on modern principles.

It is true that Hobbes stressed the fragility of justice and of civil society. The urgency of founding politics on a more stable basis gave him a strong reason to limit the

scope of politics to the basic goal of peace. This gave Hobbes another reason to limit the scope of punishment to deterrence, which greatly promotes peace, and to exclude retribution, which does not promote peace and works at cross-purposes with deterrence. But having accepted Hobbes's important insight into the fragility of politics, is it not still possible to make some room in modern criminal justice for retribution, even if this should require sacrificing some of its capacity to deter, without compromising the stability of the civil order? Considering the overwhelming power of the state in comparison with the weakness of the individual criminal, it would not be excessively dangerous to suggest that we take into account what the criminal deserves in addition to what society needs for its protection. In fact, could we not find in retributive punishment the solution to the uniquely modern problem of the conflict between society's right to punish and the natural right of the criminal? Is it not possible that corporal and carceral punishment, and perhaps even capital punishment, might be justified as what a criminal might deserve as a naturally free being and the bearer of rights? This line of argument was eventually pursued by Kant, who rejected Hobbes's view of the proper aim of punishment in order to complete the task Hobbes had set for himself: to articulate a moral theory of punishment grounded in modern principles.

Chapter 3: Kant's Attempt to Rehabilitate Retribution

Kant was an enthusiastic advocate for the modern project, well under way in the 18th century, of liberating man from his spiritual subjection to tradition and to the wills of other men. Like Hobbes, Kant believed that the two most important external conditions necessary for this spiritual liberation were the public dissemination of philosophy, or science, and absolute sovereignty. Only in the midst of a lawful and peaceful civil order protected by an absolute sovereign could the seeds of science and the arts bear the fruit of progress. Yet Kant had also inherited Rousseau's critical stance toward the ills of the modern age. The emancipation of the arts and sciences on the one hand, and the increase of trade and commerce on the other, had brought to European peoples the means to achieve unprecedented levels of well-being. But precisely those who had cultivated reason and luxury the most were, paradoxically, becoming more and more estranged from true contentment, Kant observed. It appeared that the progress of science and the arts multiplied needs more rapidly than it could create the means to satisfy them. Kant believed that this world-historical phenomenon was one of the signs which pointed to the deeper truth of the existence of a higher good than material well-being—namely, morality—to which the pursuit of material well-being ought to be subordinated. This major difference between Kant and Hobbes would profoundly shape the way in which each approached questions of political right, including the question of punishment.

Both Kant and Hobbes saw that as long as the natural right to self-preservation was not abandoned in the act of constituting civil society there would always be the

potential for a collision between this right and civil society's right to enforce its laws. Kant addressed this problem along the lines of Rousseau by arguing that the total alienation of natural rights—or rather, of man's "wild freedom"—was a necessary condition of the social contract (*MM* 6:315-16).¹ Hobbes had refused to avail himself of this solution (though he had briefly considered it) because he saw that it was incompatible with his purpose of establishing a doctrine of absolute sovereignty on the basis of the fear of violent death. Kant, on the other hand, had no such reservations. For Kant, civil society was not desirable merely as a means for protecting the lives of its members—or any private interests for that matter—but for its own sake, as the only political arrangement in which human relations could be governed by justice, where justice meant the totality of the universal principles of reason for governing men's interactions, as distinguished from the individual's natural right to self-preservation.

Hobbes's attempt to accommodate the collision of rights within his theory of punishment eventually became its undoing. As we saw in the previous chapter, Beccaria was able to appeal successfully to the humanity of his contemporaries—that is to say, their aversion towards a punishing power that did little more than declare war on individuals judged to be dangerous to civil society. Beccaria and his followers—the utilitarians—argued for their own theory of punishment informed by the principle of the greatest happiness of the greatest number, by which they claimed to have solved the problem of clashing rights without significantly restricting the sovereign's punishing

¹ *Metaphysics of Morals* (henceforward *MM*). In all references to Kant's works, the pagination of the standard German Academy edition will be used. Other works by Kant frequently cited in the text are referred to by the following abbreviations: *Critique of Judgment* (*CJ*); *Critique of Practical Reason* (*CPr*); *Groundwork of the Metaphysics of Morals* (*Gr*); and *Lectures on Ethics* (*LE*).

power. As we saw, however, Beccaria failed to demonstrate a clear connection between the *practice* he advocated—i.e., effective deterrence—and the modern principles on which he hoped to ground that practice. Most importantly, Beccaria was unsuccessful in his attempt to show how civil society’s right could be thought to supersede the criminal’s right to self-preservation, and so despite his best efforts Beccaria could not make any advance beyond the problem he encountered in Hobbes. In short, Beccaria’s rejection of Hobbes’s solution to the modern problem of punishment turned out to be more effective than the solution he proposed as its replacement.

Kant rightly regarded Beccaria’s solution to the modern problem of punishment as motivated by noble sentiments but as not fully and consistently supported by modern principles (cf. *MM* 6:334-35). On one particular point, however, Kant did agree with Beccaria; he acknowledged that the right to punish must not be derived from the state of nature, as it had been in Hobbes. Whereas Beccaria remained hampered by the idea of inalienable natural rights, Kant’s Rousseauian view of the social contract—as requiring the total alienation of natural rights to the sovereign “general will”—made it possible for him to try to derive the right to punish directly from the social contract (although, as we shall see, he did not in fact proceed in this way). Kant’s account of the right to punish revolutionized modern penal thought in a way that continues to be felt today, so much so that I believe it would not be misleading to suggest that the contemporary punishment debate continues to be informed by the fundamental dispute between Hobbes and Kant. The dimensions of this dispute will become clearer by the end of the present chapter.

Kant's contribution to modern punishment theory can be summarized as the attempt to revive retribution and to reestablish it on modern principles. The difficulty involved in this endeavor should be apparent. To Western societies, which are deeply influenced by Hobbes, what seems most objectionable about retributive punishment will be more immediately obvious than what is attractive about it. Because of its backward-looking character, retribution seems to conflict with our basic conviction that justice must benefit and not harm. From this point of view, Hobbes's claim that retribution is indistinguishable from a thoughtless desire for revenge reads like a truism. And yet, in spite of Hobbes's influence, there nevertheless remains something attractive about retribution. In the first place, it is difficult to dispense with certain commonsense moral notions—such as desert, guilt, innocence, and so on—that seem to presuppose retribution in some form, and all of Hobbes's reductionist attempts to redefine these notions appear to be little more than question-begging on his part. Second, Hobbes's own view exhibits some very traditional features, as we saw in previous chapters—especially his teaching about justice as a moral virtue transcending self-interest. At that time, we were led to wonder whether Hobbes was right when he concluded that retribution is incompatible with modern principles. Together, these two objections to Hobbes's condemnation of retribution constitute Kant's starting point. With these things in mind, let us now turn to examine Kant's theory of punishment.

Punishment and the State: Retribution as a Moral Duty to Others in Civil Society

According to Kant, in a state without positive law, which following Hobbes he calls the “state of nature,” each is completely free to act according to his own notions of what is right and good. For Kant, such a condition is necessarily violent and immoral:

It is not experience from which we learn of human beings’ maxim of violence and of their malevolent tendency to attack one another before external legislation endowed with power appears. It is therefore not some fact that makes coercion through public law necessary. On the contrary, however well disposed and law-abiding men might be, it still lies *a priori* in the rational idea of such a condition (one that is not rightful) that before a public lawful condition is established individual human beings, peoples, and states can never be secure against violence from one another . . . [which is] a state *devoid of justice* (*status iustitia vacuus*), in which when rights are in *dispute* (*ius controversum*), there would be no judge competent to render a verdict having rightful force. (*MM* 6:312)

Because this lawless condition is devoid of justice, men must resolve to escape it by constituting civil society. In civil society, the sovereign state adjudicates by its supreme authority any and all disputes among individuals regarding what is rightfully theirs. Since the civil order is the sole bulwark against the lawless state of nature, and thus the fundamental condition of all right, preserving it from dissolution becomes the first duty of the state.

Kant follows Hobbes in yet another important way by defining the scope of justice as limited to the preservation of external freedom. Kant differentiates between

ethical laws on the one hand—which make certain actions (such as the duty to honor one’s parents) duties and also make duty itself the incentive for performing these actions—and juridical laws on the other hand—which make certain actions (such as the duty to fulfill one’s contracts) duties but do not require that duty should be their incentive (*MM* 6:219). Both types of law are universal dictates of reason, and thus ought to be observed willingly by all human beings, but only juridical laws—including the laws we normally refer to as public laws—can be imposed by external legislation and enforced (*MM* 6:214, 218 ff.; 311). Kant defines juridical laws—and thus public legislation—as being concerned exclusively with the equal external relations, or freedom, of human beings and not with their ends, however noble or base those ends might be (*MM* 6:230).

Initially, it may appear to be perfectly rational for the state to seek to discharge its duty to preserve civil order simply by deterring would-be criminals by threat of force. This would involve maximizing the deterrent effect of punishment by calibrating penalties to offset any advantages that may be gained through crime. In most modern theories of the state prior to Kant’s, the need to stave off the lawless state of nature had been regarded as sufficient to justify an essentially deterrent penal system.² It comes as somewhat of a surprise, then, that although Kant’s view of the state of nature agrees in many respects with that of Hobbes or Locke, in his own account of punishment Kant emphatically asserts that the principle of punishment must be retribution rather than deterrence.

² See, again, Hobbes, *Leviathan*, ch. xxviii (beginning); Locke, *Second Treatise of Government*, §§ 7-13 and 87-88.

When, in the Doctrine of Right of the *Metaphysics of Morals*, Kant undertakes to explicate the right to punish, he says that “only the *law of retribution* . . . 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 . . .” (*MM* 6:332; emphasis in original). But what exactly is the principle of retribution, as dictated by strict justice? “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 . . . [;] whatever undeserved evil you inflict upon another within the people, that you inflict upon yourself; if you steal from him, you steal from yourself; if you strike him, you strike yourself; if you kill him, you kill yourself.” Punishment “must be inflicted upon [the criminal] only *because he has committed a crime*,

For a man 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 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 principle of punishment is a categorical imperative . . . (*MM* 6:331; emphases in original)

As the quoted passage makes clear, retributive punishment is made necessary, in Kant’s view, by a requirement of morality that human beings be treated also as ends, rather than as mere means to the purposes of others. Before we go any further in elaborating the

state's role in exacting retribution, it would help to review briefly why, for Kant, human beings must be treated as ends in themselves.

For Kant, politics is essentially subordinate to morality. While Kant agreed with earlier modern thinkers about the anarchic and violent character of the state of nature, he disagreed with the major conclusion they drew from this reflection. Kant denied that our admittedly strong natural aversion to violence and insecurity should be adopted as the fundamental moral standard. The far more important reason why men ought to leave the state of nature is that while in that state they cannot have dignity. Life, property, and happiness can all be good things, but without dignity the goodness of these desirable objects is substantially undermined. Kant argued that all these goods are therefore conditional, and that the only unconditional good that we can imagine is the morally good will, which performs duty for duty's sake (*Gr* 4:393). Since every human being possesses such a will, at least potentially, each of us must be treated as an ultimate end in itself. Accordingly, one of Kant's formulations of the categorical imperative—the supreme principle of morality—is “act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end” (*Gr* 4:429). By identifying retribution as (1) a categorical imperative and (2) as the principle of punishment, and (3) by explicitly connecting it with the formula of the end in itself, Kant makes clear that he regards retribution as a *moral* duty of the state to its subjects. Thus I believe interpreters of Kant who understand him as

having a theory of punishment grounded exclusively in political right are mistaken,³ although, as we shall see presently, difficulties of interpretation make this matter complex.

Kant's discussion of punishment in his *Metaphysics of Morals* is notoriously difficult to interpret as a result of its obscure formulations, partial explanations, and apparently self-contradictory statements. So far, I have presented Kant only as a straightforward retributivist, and yet there are several features of the Doctrine of Right that might lead the reader to believe that Kant had very serious reservations about retribution and that his support for it was qualified in important ways. Most fundamentally, Kant's statements in favor of retribution might be interpreted as meaning only that criminal guilt makes the offender conscious of his *liability* to suffer punishment, rather than implying that government should punish criminals *for the sake* of retribution and that it has an absolute duty to do so in every case. As we have already seen, Kant maintains that human beings must resolve to avoid the state of nature at all costs, and that once civil society is constituted it needs to use coercion to prevent a return to the state of nature. It may be argued that it is already implicit in the state's function as the enforcer of the civil order that it is authorized to punish in order to deter, not to pay back evil with evil. If deterrence is the reason why we punish, we would have no absolute moral obligation to punish *every* criminal to the full extent of his desert. Punishment would then only be just when and insofar as it has some deterrent effect. From this point of view, the principle of retribution might be better understood as a moral restriction, or side-constraint, on the deterrent function of punishment, forestalling potentially immoral

³ See, for example, Ernest Weinrib, "Law as Idea of Reason," 33-37.

excesses such as the punishment of the innocent. Before the state can punish an individual for the sake of deterrence “he must previously have been found *punishable*”—that is, deserving of harm.⁴

There is some further textual evidence for this reading of Kant. In two passages in the *Metaphysics of Morals*, Kant gives examples of justifiable exceptions to the law of retribution which may be taken as suggesting that punishment fundamentally aims at deterrence.⁵ In one of these passages, Kant relates a hypothetical case in which a shipwrecked and drowning man saves himself by pushing another shipwrecked man off a plank on which the latter had been keeping himself afloat (*MM* 6:235-36). Kant comments that a court should not punish the first man for murder, since “the punishment threatened by the law could not be greater than the loss of his own life” by drowning. Here, Kant may be taken to imply that since a penal law against murder would be ineffectual in this instance—as it could have no deterrent effect—the purpose of all penal laws, as such, must be deterrence. The same significance might be attributed to another passage, where Kant gives two more examples of crimes to which the principle of retribution ought not to be applied, because such punishment would have no deterrent effect. Kant says that a soldier who kills his opponent in a duel to defend his own honor, and a mother who kills her illegitimate child for fear of public shame, ought both to be exempted from capital punishment, since in both cases the law against murder conflicts with an overriding incentive of honor created by a “barbarous and undeveloped” custom

⁴ See Thomas Hill, Jr., “Kant on Wrongdoing, Desert, and Punishment,” 409. See also Sharon Byrd’s “Kant’s Theory of Punishment: Deterrence in its Threat, Retribution in its Execution”; and Don Scheid’s “Kant’s Retributivism.”

⁵ See Mark Tunick, “Is Kant a Retributivist?,” 60-78.

(*MM* 6:336-37). Although in both cases a killing has been committed, the penalty of death (as determined by the principle of retribution) would not be appropriate, since in both cases the incentive of honor is strong enough to overcome the fear of death to make the punishment an ineffective deterrent. As in the case of the drowning man, the suggestion would seem to be that retribution is subordinate to deterrence, and that the latter is therefore the real aim of punishment.⁶

This interpretation does not seem to me to be the most plausible reading of Kant, nor do I think that in the two passages I have mentioned Kant is going back on his earlier, unambiguous statements that “*only* the law of retribution . . . can specify definitely the quality and the quantity of punishment” and that such punishment is “a *categorical imperative*” (my emphases). It is much more likely that Kant understood the examples he gave of exceptions to the law against murder to be just that: exceptions to the law made necessary by some exigency, and implying nothing about the purpose of the law itself. Exceptions, as exceptions, do not determine the rule. Therefore, I do not think that Kant’s reference to the failure of deterrence in these cases as a reason to excuse killings from retributive punishment is meant to suggest that punishment *in general* ought to aim at deterrence. Kant himself says, in the context of the example of the drowning man, that “necessity has no law,” thereby seeming to imply that killing for self-preservation in such circumstances could not even be considered as murder in the strict sense (*MM* 6:236). In the other passage, Kant says the same thing explicitly about the acts of the duelist and the mother of the illegitimate baby: “it seems that in these two cases people find themselves

⁶ See also Tunick’s discussion of the sovereign’s right of clemency at *ibid.*, 63-64.

in the state of nature, and that these acts of *killing (homocidium)* . . . would then not even have to be called murder (*homocidium dolosum*)” (*MM* 6:336). But a more obvious reason to reject this alternative reading of Kant is simply that Kant himself never says explicitly that the principal aim of punishment ought to be deterrence—at least not in his thematic treatment of punishment in the *Metaphysics of Morals*—whereas he does state explicitly that retribution is the principle of punishment.⁷ Now that I have established that Kant did indeed believe that retribution is a moral duty of the state to its subjects, I will try to explain more fully *why* he believed this to be true.

Deterrence and the Principle of Reciprocal Coercion

Because of his status as an end in itself, man is the highest being in nature. In relation to the rest of nature man is an absolute master who possesses an unlimited right to use, transform, and even destroy non-rational beings according to his will (*MM* 6:246-7; *CJ* 5:429-31). In practice, however, there are many equal rational beings coexisting in this natural state, each with an equally rightful claim to unlimited external freedom. This inevitably brings human beings into conflict over external objects—that is, over land, food, and other resources. As a result of their moral equality, men in the state of nature are required by reason to observe the law of external reciprocity, according to which any

⁷ There is another passage that is cited by some Kant scholars as showing him to be espousing what is essentially a deterrence theory of punishment. In his posthumously published *Lectures on Ethics*, Kant says that “[a]ll punishments imposed by sovereigns and governments are pragmatic; they are designed either to correct or to make an example” (*LE*, “Reward and Punishment,” 55). They are not, in other words, designed to punish wrongdoers according to their desert. I can find no plausible interpretation of this passage that would reconcile it with the interpretation of Kant I advance here. Nevertheless, if, in the final analysis, there remains a real discrepancy between the two texts, more weight should be given to the *Metaphysics of Morals*, which was published in Kant’s lifetime, as representing his mature view, than to his lectures, which were published without his supervision.

action is right as long as it can coexist with everyone else's equal freedom (*MM* 6:230-32). Moreover, every person is authorized to enforce this law in the otherwise lawless state of nature, and each does so without impinging on anyone else's right, since he does so "as a hindering of a hindrance to freedom." This universal authorization of reciprocal coercion then becomes the source of all right (*MM* 6:231). But since in the state of nature, where each is judge in his own case, there will be disagreements about rights with no recourse but to violence, all must resolve to leave this state and enter a lawful political condition where conflicts can be arbitrated by the common authority of a sovereign state in accordance with right, and where those who would defy the authoritative arbitration could be compelled to submit to it by force (see again *MM* 6:307 ff.).

Necessity and prudence require the state to use deterrent threats in order to discharge its duty of protecting the civil order from dissolving back into the state of nature, but the logic of deterrence would lead it to use individuals as mere means—which would be a morally unacceptable consequence for Kant—in the following way. The principle of reciprocity authorizes "hindering of hindrances to freedom," and so permits the state (1) to *resist* encroachments by individuals on one another's freedom and (2) to *restore* unlawfully dispossessed property. But, at the same time, it does not seem to me that this principle, as stated, could justify the preemptive coercion of individuals by the state in order to prevent them from *attempting* to commit crime in the future, as this would deprive them of their rightful freedom. I can call the police to come to my aid to prevent you from robbing me or, if you succeeded in robbing me, the principle of reciprocal coercion could authorize the state to restore my property to me, but it cannot

authorize the state to incarcerate you so as to prevent you from trying to rob me again, or to make an example of you to prevent others from doing so. The latter could not be understood as a hindrance of a hindrance to freedom because, in this case, the initial hindrance has not yet occurred, and it would be unlike Kant to anticipate the future choices of a free will. Kant does indeed argue that a thief owes a debt to society as a whole, having made all property insecure, and that for this reason he may be punished with incarceration. But this backward-looking argument is already an argument from *retribution*, not from *reciprocity* (see *MM* 6:332-3 and context).⁸ If the state determines the quality and quantity of penalties with a view to deterrence alone, and without having those penalties reflect the specific gravity of each crime—i.e., its quality and quantity—then punishment will go beyond what is due to a rational agent, which is the moral consequence of his choices, and thus will result in treating the criminal contrary to what is required by his dignity and status as a rational agent.

But, to repeat, the state still has a need to preserve the civil order, and this cannot be done effectively without the use of deterrence. It seems to me—and in this I depart from most Kant scholars—that from Kant’s point of view the only way of overcoming this dilemma was for him to try to ground the state’s use of deterrence in some other purpose that would at the same time allow the state to treat criminals as ends in themselves. According to Kant, that purpose is retribution, which fits punishment to the crime and so reflects the criminal’s choice (i.e., to commit that very crime), thus respecting or affirming his dignity as a rational agent. Kant fully expected that retributive

⁸ But compare Susan Shell’s account, in which she attempts to derive the right to punish retributively from the principle of reciprocity itself, in “Kant on Punishment,” 116-24.

punishment would in practice have a *deterrent* effect as well, which would allow for civil society to preserve itself. But since this would not be the principle, or the primary aim, of punishment, the state would avoid degrading criminals by making their suffering a means to the security of other men.

According to Kant, we may be held responsible not only for the external aspect of our actions, but also for the maxims that serve as grounds for our choices (*MM* 6:225). In plain terms, this means we may be held accountable for choosing justice or injustice. We may impute to a murderer not only the death of his victim, but also his decision to act according to the immoral maxim that one's private hatred (for example) should count for more than the lives of other rational beings. No matter how wicked and deplorable the criminal's reasons for acting may have been, he remains a rational, and thus autonomous, being whose immoral but free choices must be respected rather than disregarded as mere mistakes or accidents. According to Kant, retribution does just this by punishing the offender "in proportion to his *inner wickedness*," and thereby acknowledges his actions as rational (*MM* 6:333; emphasis in original).⁹ Paradoxical as it may sound—and perhaps it is only paradoxical because of the influence of Hobbes on our habitual beliefs—to punish a criminal retributively is, according to Kant, to pay him the highest compliment, whereas to let him off the hook is to condemn him as childish, irresponsible, and

⁹ In the context, Kant anticipates the objection that equal retribution "is not possible in terms of the letter" in every single case. No one can take an eye for an eye from a blind man. Kant argues, however, that it is possible to remain true to the spirit of the principle, if not always to its letter. In his example, Kant suggests that when an innocent person has been insulted, the punishment of the offender can satisfy the demand of retributive justice, even if it does not take the exact same form as the original insult, as long as the offender is made to feel shame in proportion to the outrage he caused his victim. *MM* 6:332-33.

something less than a full human being.¹⁰ One might say that, for Kant, retributive punishment forces the criminal to be free, to borrow a phrase from Rousseau. It is partly for this reason that Kant regarded retributive punishment not just as a right, but as a duty or categorical imperative. He illustrates this point most dramatically and unambiguously when, in commenting on the hypothetical example of a society about to dissolve by the consent of its members—i.e., when punishment can no longer have a deterrent effect—he says that in such circumstances the last murderer remaining in prison must still be executed “so that each has done to him what his deeds deserve” (*MM* 6:333).

Some Difficulties with Kant’s Position

There are, however, still serious problems with the position I have attributed to Kant, problems that have led some scholars to reject his account of punishment as incoherent.¹¹ In the first place, as I have already noted, Kant draws a sharp distinction between juridical and ethical laws. Ethical laws must be obeyed for the right reasons—i.e., for duty’s sake—and the performance of these laws cannot be externally coerced. On the other hand, juridical laws, which can be enforced, constrain only our outward behavior and are indifferent to the incentives which move us. How, then, can criminals be punished for their immoral motives if no one can be forced to act morally? Second, there is the fact that Kant says that while God can see into our souls and know our true

¹⁰ This insight has been expressed more recently, though not as systematically as in Kant, by Herbert Morris, “Persons and Punishment”; and by Jeffrie Murphy, “Marxism and Retribution.”

¹¹ See, e.g., Jeffrie Murphy, “Does Kant Have a Theory of Punishment?”

motivations, human beings have no such insight.¹² How, then, could the state retributively punish a criminal “in proportion to his inner wickedness” if it cannot exactly know his true motivations?

I will address the second difficulty presently and postpone for now the question of the first difficulty. One particular passage in the account of punishment in the Doctrine of Right suggests that Kant recognized and wrestled with the tension between the moral need to punish criminals as they deserve and the difficulty we have discerning one another’s true motivations. Kant seems there to advance a tentative solution to the dilemma. He does so by trying to show that in the case of capital crimes the “fitting of punishment to the crime” is guaranteed by “imposing the death sentence in accordance with the strict law of retribution,” since “only by this is a sentence of death pronounced on every criminal in proportion to his *inner wickedness*” (*MM* 6:333 ff.; emphasis in original). To explain his point by way of illustration Kant gives the example of two rebels: one who acts from the honorable (though mistaken) motive of wishing to depose a government he regards as illegitimate, and another who acts only for the sake of his private gain. Kant argues that both would receive what they deserve if they were sentenced to die. “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 [i.e., valuing honor more than life] and the scoundrel severely in terms of his [i.e., valuing life more than honor]” (*MM* 6:334). Thus, there would be no need for the state to inquire into

¹² See *Critique of Pure Reason*, A 551-52/B 579-80 (475).

the two criminals' actual motivations; whatever their motivations happened to be, a sentence of death (as dictated by the principle of retribution) is the fitting punishment!

This would indeed be an ingenious solution to the problem if this rule could be shown to hold generally in all cases of possible crimes and motivations. Unfortunately, it is very unlikely that it *would* hold generally, and thus Kant's solution is partial at best. At any rate, what is important is that this passage clearly indicates that Kant believed that punishment must reflect in some way the *ends* of men's choices. That is to say, punishment must requite deeds in terms of what gives them their moral worth—namely, their motivations.¹³ One of the rebels deserved a milder punishment because he was motivated by honor, while the other deserved a more severe one because he was motivated by profit. In addition, it must be added that, for Kant, not punishing for fear of making an error of judgment about motivations may have worse moral consequences than punishing and making the occasional mistake, since such an error of judgment would at most cause undue physical harm, while refusing to punish retributively altogether would be equivalent to denying the criminal's rationality, which would be dehumanizing.¹⁴

The argument thus far has traced the dialectical stages in the development of Kant's account of punishment. If correct, it shows that Kant regarded retribution as the principle of punishment and its deterrent effects as incidental, albeit desirable, consequences. As we have seen, the principle of reciprocal coercion (the right to hinder

¹³ I must therefore disagree with Herbert and with Fleischacker, both of whom attempt to give some other meaning to Kant's use of the term "desert." See Gary Herbert, "Immanuel Kant: Punishment and the Political Preconditions of Moral Existence," 67-72; Samuel Fleischacker, "Kant's Theory of Punishment," 202-5.

¹⁴ See also Herbert, "Immanuel Kant: Punishment and the Political Preconditions of Moral Existence," 67-68.

hindrances to freedom), which is the source of all right, justifies only the use of preventive and restorative measures by the state but does not justify deterrence, which is nevertheless necessary for the state to preserve itself. Strictly speaking, then, the right of reciprocal coercion creates the *need* for deterrence, but deterrence is made possible only as a byproduct of retributive punishment. Retribution, in turn, is a moral imperative to treat criminals as ends in themselves.

Nevertheless, we are still tempted to ask whether the preceding is not just a specious distinction employed by Kant to give a morally attractive semblance to what is in fact a penal theory essentially aiming at deterrence. If the *raison d'être* of retributive punishment is to make possible the exercise of deterrence by the state, then is it not plain that retribution exists for the sake of deterrence and not vice-versa? If so, then perhaps the logic of Kant's theory of punishment does *not* dictate an absolute duty to punish the guilty, particularly in those cases in which punishing would not contribute to deterrence (e.g., in a society about to dissolve itself the last murderer would then *not* have to be executed).¹⁵ This would be perfectly compatible with the kind of side-constraint or limiting retributivism I sketched above.

In suggesting this criticism I do not mean to imply that it is necessarily correct. It is important to acknowledge, however, that it does point to a fundamental problem that haunts Kant's political teaching as a whole, which can be stated as follows. If right is to be grounded in human freedom, then no part of it can be alloyed with men's natural needs, since the realm of nature is the realm of necessity, according to Kant. The

¹⁵ Cf. Tunick, "Is Kant a Retributivist?," 77-78.

principles of right must therefore be derived entirely from a priori reason. And yet, from time to time, Kant appears to base his arguments on empirical considerations—although he never regards himself as doing so. Much turns on the question of whether or not Kant remains consistent in this respect. Arguably, his position on punishment is one instance of this. Kant seems to accept that the empirical fact of criminality makes deterrence necessary, and then goes on to argue that this creates a moral requirement for punishment according to an a priori principle—i.e., retribution. He then draws the conclusion that, since a priori principles are by their nature unconditional, retributive punishment is to be regarded as a categorical or absolute duty. The obvious question raised by this argument is to what extent Kant’s accommodation of the empirical need for deterrence compromises the purity, and with it the absolute character, of retributive punishment.

In order to answer this question, it would help if we formulate it in a slightly different way: Can retributive punishment be shown to be something noble, or intrinsically good for its own sake? If it could be shown to be noble, it could then plausibly be argued that retribution is morally desirable in all cases, even when it would not contribute in any way to deterrence. On the other hand, if it could not be shown to be noble in its own right, then the whole institution of punishment would have to be understood as only a necessary remedy to the problem of the state of nature, and thus the principle of retribution would have to be conceived as, at most, a side-constraint on the deterrence function of punishment. We will have more to say about this important question later in this chapter.

We have now traced the arguments by which Kant expresses and defends his view that deterrence is insufficient to justify punishment on moral grounds and that retribution must therefore be the principle of punishment in a lawful civil order. The cogency of these arguments depends on whether or not Kant's claim that retribution preserves intact the inner dignity of the offender, as a rational being, is true. But, so far, we have only accepted this as a formulaic premise. What proof is there that the offender himself really does experience his punishment as dignifying and morally elevating? Kant's account of retribution in the *Metaphysics of Morals* is less than fully satisfying because it does not show how its formulaic pronouncements connect to ordinary moral experience, and so it has to be put together with what he has to say about this matter elsewhere. Finally, if we take seriously Kant's distinction between juridical and ethical laws, then how can we punish criminals for their immoral motivations if no one can be coerced to act morally? The next two sections attempt to address these questions.

Retribution and the Conscience

What separates human beings from the rest of nature, according to Kant, is our capacity to act as beings subject to universal legislation. Since this capacity is innate to humanity, each may hold himself in the highest esteem and may claim a right to be respected by others on an equal footing, regardless of the position in which he has been placed by nature or society (*MM* 6:434-436). Although Kant tends to emphasize this egalitarian basis of personal worth in his works, he also acknowledges another source of worth for which we can be held in esteem—namely, the goodness of our actions. An

example of the latter may be found in the familiar case of Kant's world-weary and reluctant philanthropist who, "no longer moved by any inclination . . . tears himself out of this deadly insensibility and does the action [i.e., helping others in need] without any inclination for the sake of duty alone." At that moment "for the first time his action has its genuine moral worth" (*Gr* 4:398-99). Although, as a human being, the reluctant philanthropist already deserves respect, his difficult and rare self-conquest elevates still higher his standing in our eyes, and in his own, and is responsible for the moral flourishing he experiences.

If innate humanity alone is not sufficient to account for the moral flourishing described in the reluctant philanthropist's example, then how exactly can we account for it? Occasionally, Kant argues that what underlies our admiration for good actions is really our respect for the moral law as it is manifested in and through good actions. But this alone does not exhaust the full meaning of the experience of moral flourishing, as Kant himself admits, and as brief reflection will show. In the first place, we know that contemplating our virtuous actions is positively enjoyable, and that this enjoyment lingers for a time after the virtuous actions have been completed—and thus presumably after the moral law has ceased actively manifesting itself through our actions. Second, we believe ourselves to be entitled to this enjoyment of our virtuous accomplishments as our proper and fitting reward. As Kant puts it in his discussion of duties of virtue, "there is a subjective principle of ethical *reward*, that is, a susceptibility to being rewarded in accordance with laws of virtue: the reward, namely, of a moral pleasure that goes beyond mere contentment with oneself . . . and that is celebrated in the saying that, through

consciousness of this pleasure, virtue is its own reward” (*MM* 6:391; emphasis in original). It is remarkable that this gloss on the saying that virtue is its own reward denies the literal interpretation of this saying. Virtue is not its own reward because virtue essentially involves a painful struggle against one’s natural inclinations (cf. *MM* 6:394, 6:379-81). Virtuous men are, however, entitled to further compensation, part of which is the moral pleasure of self-esteem. The case of self-reproach is similar to that of self-esteem. When we judge ourselves to have done wrong in some way, we believe ourselves to deserve the feeling of regret and self-reproach that we suffer on account of it. Thus pride and self-reproach, both essential to the inner life of morality, may be regarded as self-imposed forms of reward and punishment.

On Kant’s view, the function of the conscience is analogous to that of a judge in another, more direct way. When the conscience passes judgment it “pronounces the sentence of *happiness* or *misery*, as the *moral results* of the deed” (*MM* 6:439 n.; my emphases). The conscience does not merely identify certain types of actions as either morally good or bad. Rather, its judgment is also a “sentence” bearing consequences: it “either acquits or declares us guilty and deserving of punishment” (*LE*, “Conscience,” 129). Nor is it sufficient that this judicial pronouncement of the conscience is appropriately followed by self-reproach and repentance, but only “if it is felt *and enforced*” (*ibid.*, 131; my emphasis). According to Kant:

The first effectual expression of this judicial verdict which has the force of law is moral repentance; the second, without which the sentence is inoperative, is action in accordance

with the judicial verdict. If it does not result in *practical endeavor* to do what is demanded for *the satisfaction* of the moral law, the conscience is but an idle conscience, and however penitent we may be the penitence is vain so long as we do not *satisfy the debt we owe* to the moral law; for even *in foro humano* a *debt* is not satisfied by penitence, but by *payment*. (Ibid.; my emphases)

The “payment” of the “debt” we owe to the moral law, as distinguished from mere “penitence,” clearly refers to punishment. Although Kant does not spell it out, his meaning seems to be that a person with a healthy conscience must be willing not only to accuse himself, but also to submit to punishment as a morally necessary consequence. If he somehow avoids punishment, then he loses his basis for self-respect and is degraded in his own eyes. “Preachers must, therefore, impress upon their hearers that, whilst they must repent for their transgressions against their duties to themselves, though they cannot remedy these, in the case of injustice done to others mere repentance is not enough: it must be followed by endeavor to remedy the injustice” (ibid.). To salvage his dignity as a moral being the offender must seek atonement by submitting himself to punishment.¹⁶

Kant’s advice to preachers to urge members of their congregations who have committed injustices to turn themselves over to their fellows so that they can be judged and punished for the sake of their own moral restoration calls to mind our earlier question of how to square retributive punishment with Kant’s distinction between juridical laws, which are enforceable, and ethical laws, which are not. This difficulty may perhaps be

¹⁶ As an illustration of what Kant may have in mind in this passage, consider the example of the titular character of Thomas Hardy’s *The Mayor of Casterbridge*.

resolved if we suppose that “enforceable” does not mean the same thing as “punishable.” The function of punishment proper, we might say, is to *redeem* wrongs according to their moral worth, whereas that of law-enforcement (which would include cops on the beat as well as any deterrent effect of punishment) is to promote compliance with the law. This distinction helps to explain why a conscientious person should willingly submit himself to the punishment that he deserves in order to undergo the kind of moral restoration that Kant describes. Yet this would still not explain why retribution should be imposed by the state as an act of justice, if the aim of justice is limited to regulating only the external relations of men. It would appear that in the case of retributive punishment Kant is attempting to extend the scope of justice by having it serve the ends of morality, and not merely the more narrow ends of external freedom. If this is indeed the case, then Kant would be contradicting the fundamental modern separation of justice from morality, which he himself explicitly adopts in his political writings.

In reflecting on Kant’s view of the conscience, in light of which retributive punishment might indeed be seen as dignifying and morally elevating even by the criminal himself, we cannot help but notice how much it owes to the Biblical tradition.¹⁷ So far, we have only dealt with Kant’s view of retribution strictly in the context of political justice; we have not yet examined his view of it in the context of divine justice. It now seems necessary to investigate the latter as well, as it has become apparent that Kant had been profoundly influenced by the Scriptures in his view of punishment, and

¹⁷ On the conscience, cf. Romans 2:15 and Proverbs 20:27; on retribution cf. Romans 2:5-12 and Proverbs 20:30.

not just by modern principles. Specifically, we must ask whether or to what extent Kant manages to reconcile these two influences within a coherent theory of punishment.

Divine Retribution and the Summum Bonum

According to Kant, when human beings are acting morally they are naturally led to make certain theological assumptions or postulates, even if they are unaware of doing so. The postulation of the existence of an omnipotent, intelligent, and just being that always acts in accordance with the laws of morality arises out of a need to resolve the great antinomy of practical reason between happiness and morality (*CPr* 5:110 ff.). As natural beings we are constituted by our needs, the satisfaction of all of which taken together is called happiness. On the other hand, as moral beings we believe that we ought to pursue virtue above all else—a good higher than happiness and choiceworthy for its own sake—regardless of whether or not it also contributes to our well-being. The complete good, or *summum bonum*, for human beings is to possess both goods at once, happiness and moral rectitude, even though in nature there seems to be no necessary correlation between the two. We do not think that by gratifying our desires we will become virtuous, nor do we have any reason to expect that by acting virtuously we will necessarily become happy. As both experience and natural science teach us, there is no necessary connection between acting well and faring well in the natural world (*CPr* 5:110-14).

Still, precisely when we are acting morally we cannot help but believe, Kant insists, that the *summum bonum* is possible and, moreover, that we can bring it about

through our own actions—for if it were otherwise then “the moral law . . . must be fantastic, directed to empty imaginary ends, and consequently inherently false” (*CPr* 5:114). It is from these facts of the active moral consciousness that Kant deduces his general principle that if “virtue and happiness are thought of as necessarily combined, so that the one cannot be assumed by a practical reason without the other belonging to it,” then this combination must be “as the connection of cause and effect” (*CPr* 5:113). The only way in which we can conceive of the connection between virtue and happiness as *causal* is if we postulate the immortality of the soul and the existence of a just and omnipotent God (*CPr* 5:114-132). On Kant’s view, the existence of God would guarantee the “effect” of happiness as a consequence of moral goodness because God could be expected to distribute well-being to each in exact proportion to his desert, “for to be in need of happiness and also worthy of it and yet not to partake of it could not be in accordance with the complete volition of an omnipotent rational being” (*CPr* 5:110). Kant claims that only on the assumption of such postulates is morality possible, since only then can we allow ourselves to hope that human beings will be rewarded and punished in accordance with their moral worthiness by an unerring judge, if not in this life then in the next.

This leads to the terrifying and awesome implication that each person will be held to account by an omniscient and all-powerful judge for every one of his actions. What’s more, while Kant’s God is a benevolent lover of mankind, he is above all a lover of justice for its own sake. For in distributing happiness in accordance with the concept of the *summum bonum* God does not will that all men should become perfectly happy, but

rather that they should enjoy as much happiness as they have deserved through their deeds.¹⁸ In his *Critique of Judgment*, Kant says that the final purpose of nature, as created by God, “can only be *man under moral laws*” (CJ 5:445; emphasis in original). But Kant continues:

I say deliberately: *under* moral laws. The final purpose of creation is not man [acting] *in accordance with* moral laws, i.e., a man whose behavior conforms to them . . . And this agrees perfectly with the judgment that human reason makes when it reflects morally on the course of the world. Even in evil we believe we perceive the traces of a wise reference to a purpose, provided we see that the wanton villain does not die until he has suffered the punishment he deserves for his misdeeds. . . . [T]he highest wisdom in the government of the world we posit in this: that the opportunity for good conduct, but the consequence of both good and bad conduct, is ordained according to moral laws. In the latter consists, properly speaking, the glory of God, and hence it is not unfitting if theologians call it the ultimate purpose of creation. (CJ 5:449, n.; emphases in original)

According to Kant, then, the ultimate end of divine providence, the *summum bonum*, is not *necessarily* men acting in complete conformity with moral laws. Rather, the end of providence is perfect justice: the proportional distribution of happiness in accordance with moral worthiness.¹⁹ In this and not in God’s unbounded benevolence consists the “highest wisdom” of his governance.

¹⁸ Cf. again Romans 2:5-12.

¹⁹ Cf. Peter Byrne, *Kant on God*, 110-17; Lewis White Beck, *Commentary on Kant’s Critique of Practical Reason*, 270-71.

This ringing endorsement of divine retribution is striking and highly significant for our understanding of Kant’s view of both divine providence and judicial punishment. Kant had argued that human judges must embody the principle of retribution as the only way of guaranteeing that the human dignity of criminals will be respected. In our discussion of judicial punishment above, I found occasion to ask whether retribution is really an absolute moral duty (a categorical imperative), or whether it is only a moral limitation or side-constraint on the state’s use of deterrence. There, I suggested that this question turned on whether or not retributive punishment is noble in its own right or, to put it differently, whether it is intrinsically good. As we can now see, retributive justice appears to Kant as nothing less than the “highest wisdom” of God’s governance of the world and a core feature of the ultimate end of practical reason, the *summum bonum*. If retributive punishment is part of divine governance, then it would seem that retributive punishment is indeed, for Kant, an intrinsic moral good.²⁰ Could this, then, be the deeper reason why Kant regarded retribution as the state’s absolute moral duty, as opposed to a limiting principle or side-constraint on the use of deterrence? If so, could it be that judicial retribution is, for Kant, a kind of analogue or extension of divine retribution?²¹

I pose this as a question because there is an obvious difference between divine governance and political rule, and so it could not be argued that judicial punishment must be retributive because divine punishment is retributive. This is not the argument I am making. Rather, I am suggesting that the fact that God punishes retributively—on Kant’s

²⁰ In the *Critique of Pure Reason*, where Kant denies that punishment would be necessary in an ideal state, he is working on the counterfactual assumption that crime does not exist. See A 316-17/B 372-74 (311-12).

²¹ This is Fleischacker’s speculation as well. See “Kant’s Theory of Punishment,” 203-6.

view—may have confirmed for him that retribution is intrinsically good or noble, as part of the noble divinely created order, and this may be what ultimately persuades Kant that to punish a man retributively is to treat him with dignity. Surely, God’s governance over human beings must respect their dignity, and if this ideal governance includes retributive punishment—indeed, if in this we discern its “highest wisdom”—then it is quite possible that Kant believed retributive punishment to be an important condition of human dignity. If Kant also saw the principle of pure deterrence as seriously threatening human dignity, then it would make sense that he should want to ground modern punishment in the principle of retribution.

What can we now say about Kant’s position on the intrinsic moral goodness of cosmic retribution, as it has come to sight? On the one hand, it resonates with our own basic understanding of one important aspect of justice. “When . . . someone who delights in annoying and vexing peace-loving folk receives at last a right good beating, it is certainly an ill, *but everyone approves of it and considers it as good in itself even if nothing further results from it*; nay, even he who gets the beating must acknowledge, in his reason, that justice has been done to him, because he sees the proportion between welfare and well-doing, which reason inevitably holds before him, here put into practice” (*CPr* 5:61; my emphasis). This is an easily recognizable expression of righteous indignation which all of us have felt at some point in our lives. On the other hand, we are still compelled to wonder why God’s (and our) sense of justice must be so strict and unforgiving towards those who commit injustice. Would it not be better from a moral point of view if strict justice were occasionally softened either by high-minded generosity

or by what Aristotle had called equity? Yet in his own discussion of equity, Kant gives no indication that he would approve of an equitable dispensation from the judge's duty to apply the principle of retribution (cf. *MM* 6:234-35), not to mention the fact that this is suggested by the stress Kant puts on the categorical character of that principle and on the need for its strict application (as in the case of the last murderer in Kant's example).

The most obvious Kantian response to the last objection would be to repeat the steps of his derivation of the principle of the *summum bonum*: practical reason requires that happiness be dependent on virtue "as the connection between cause and effect," and this makes retribution (in the form of reward and punishment) an absolute duty. In Kant's view, men cannot be denied this particular kind of justice by the state, lest the integrity of justice as a whole should be compromised (cf. *MM* 6:331-32). But, as some of Kant's critics have pointed out, several problems with this connection between virtue and happiness make it difficult to accept. One problem is that Kant does not adequately show how the element of happiness in the *summum bonum* might be derived from the categorical imperative if, as he maintains, the categorical imperative is a purely formal principle. As a purely formal principle, the categorical imperative cannot determine the content or end of moral action, and thus it is not completely clear why Kant thinks that happiness—let alone happiness distributed in proportion with moral goodness—must be an end of a priori practical reason.²²

Another problem is that Kant's position seems to blur together two distinct issues: *moral worth* as an unconditioned good on the one hand, and *moral worthiness* as the

²² See Beck, *Commentary*, 242-45. Cf. *Gr* 4:71-72.

condition of entitlement to happiness on the other.²³ It makes sense that what is morally right should always be chosen over mere physical contentment if or when the two are in conflict. But this is not by itself sufficient to substantiate the claim that happiness should be distributed *in exact proportion* to virtue. If happiness and virtue are both goods—albeit of unequal rank—why should the first be dependent on the second, and why should this dependence take the shape of geometric proportionality? Why should we conceive of their relation as proportional to one another rather than cumulative? It is true that Kant maintains throughout his moral writings that happiness is a “conditional” good, but he seems to do so merely in order to emphasize that the good will is the only *unconditional* good. This alone is hardly sufficient to give Kant the right to conclude that the goodness of happiness ought to be conditioned *by virtue*.

That happiness ought to be dependent on virtue through the agency of God, as supreme dispenser of cosmic retribution, had been judged by Kant to follow from the immediate facts of the active moral consciousness, and the most relevant of these facts was said to be our hope that all men should attain just as much happiness as they deserve. But is the moral consciousness as unequivocal about this as it appears from Kant’s presentation? Contrary to Kant, doesn’t the moral consciousness itself approve of the occasional softening or suspension of strict justice out of generosity, benevolence, or equity, as we have just pointed out? Is it not possible that the moral consciousness itself is much more divided than Kant admits as to the goodness of strict retribution?

²³ See Susan Shell, *The Rights of Reason*, 94. For yet another criticism of Kant’s concept of the *summum bonum*, see William A. Galston, *Kant and the Problem of History*, 254-55.

The same criticism can be made in a somewhat different way. One major challenge posed by Kant's thought to all prior moral philosophy is his repudiation of all attempts to understand morality by means of theoretical or speculative reason.²⁴ Throughout his works, Kant maintains that basic moral concepts (e.g., the origin of moral laws, the freedom of the will, the existence of God, etc.) can only be understood from the point of view of the moral consciousness as an independent source of knowledge. Accordingly, Kant denies the need for any theoretical proof of the existence of God and the immortality of the soul, and contends that it is sufficient to establish these as postulates required by our finite moral consciousness. Perhaps, then, we should likewise try to understand the dependence of happiness on virtue—which is presupposed by the retributive outlook—as a necessary postulate of our finite moral consciousness. No further „metaphysical’ reasons would then be required to explain why we should approve of this.

But this way of understanding Kant does not seem to me to escape the criticism I have been pressing. It is true that Kant does argue for the independence of morality from speculative reason, and yet he never ceases to insist that morality is still grounded in (practical) *reason*. If we take this claim seriously as an expression of Kant's rationalism, then we must expect his moral concepts to follow perspicuously from the basic facts of the moral consciousness, even if we cannot demand that we should be able to prove the actual existence in nature of the things to which these concepts refer. Kant's explanations for the postulates of the existence of God and the immortality of the soul seem to me to

²⁴ For the following interpretation of Kant I am indebted to Emil Fackenheim's "Kant's Philosophy of Religion."

meet this test, at least provisionally. If, as Kant claims, our moral consciousness informs us that we have an absolute duty to conform to the moral law, and since we know perfect moral goodness to be impossible for us as long as we are finite beings endowed with natural inclinations, it follows that we can fulfill our duty to be morally perfect only in an afterlife, that is, only if our soul is immortal. As for the postulate of God's existence, although Kant derives it with reference to the *summum bonum* (which I am now questioning) there is also an alternative derivation which is equally effective but more straightforward. Since we are not the authors of nature, we cannot assume that the laws of nature can coexist with the laws of morality—i.e., we cannot know whether the deterministic laws of nature allow for the possibility of moral agency; and yet our moral consciousness compels us to believe, according to Kant, that moral agency *is* possible for us, and so we must postulate the existence of an omnipotent author of nature who guarantees its possibility. These postulates follow straightforwardly from the facts or needs of the moral consciousness, as Kant describes them. But if the proportional distribution of happiness in accordance with virtue is to be understood as another postulate of practical reason, is it as clear and intelligible as the postulates of the existence of God and of the immortality of the soul?

Although we can agree with Kant that the moral consciousness tends to approve of the proportional punishment of wrongs *if* they should occur, I do not believe that it can support the claim that wrongdoing and retribution ought to be part of the *ideal* order of the world, as governed by an omnipotent and just God. Indeed, Kant himself seems to

admit, at the very end of the *Metaphysics of Morals*, that divine retribution is a great mystery that defeats human reason:

For, in view of the eventual multitude of criminals who keep the register of their guilt running on and on, punitive justice would make the *end* of creation consist not in the creator's *love* (as one must yet think it to be) but rather in the strict observance of His *right* [T]his seems to contradict principles of practical reason, by which the creation of a world must have been omitted if it would have produced a result so contrary to the intention of its author, which can have only love for its basis. (*MM* 6:490-91)²⁵

What we have just seen gives us further reason to wonder to what extent Kant continues to build his theory of punishment upon modern principles, and to what extent he departs from those principles by allowing himself to be guided by the premodern insights he adopts from the Bible. As Kant admits, unassisted reason leads us to wonder whether a retributive God would really be a demonstrably benevolent God. Must we then, in the final analysis, abandon reason and accept divine retribution as part of the awesome mystery that is divine justice? If this is ultimately Kant's position, then his retributivism would seem to be less an expression of his rationalism than of his Biblical faith.

It appears to me that Kant himself *believed* that the Biblical view of retribution is compatible with modern principles. At one point, Kant even tries makes a brief attempt to show that the kind of strategy he used all along for deriving the law of retribution was the same old contractarian strategy used by Hobbes and Beccaria (see *MM* 6:334-35).

²⁵ Cf. *Religion within the Bounds of Reason Alone*, 6:5-6.

According to Kant's version of the contractarian strategy, men give up all their natural rights when they conclude the social contract and form the sovereign "general will." As a member of the sovereign general will, each person is a legislator who subjects himself to public laws, including penal laws, and judges himself to be punishable under those laws if he should ever be found guilty of a crime. In this way, the criminal authorizes his own punishment. But what Kant does not explain in this account, and what he cannot explain on modern principles, is why the penal laws that are authorized by each co-legislator must be *retributive* penal laws. According to the modern view, the purpose of justice is to protect external freedom—it is not to fulfill the ends of morality. Consequently, nothing in the social contract itself requires retribution, and so it seems necessary to conclude that the real source of Kant's retributivism lies outside of modern principles—perhaps in Scripture, as I have suggested. Inasmuch as these problems in Kant's account of retributive punishment remain unresolved, it seems we must judge it to be inadequate in the final analysis, at least from the point of view of modern political rationalism.

Broadening the Scope of the Question of Punishment

The foregoing has been an attempt to explore Kant's view of punishment in order to help us to resolve our modern perplexity on this subject. Working through Kant's sensitive reflections about the moral consciousness has deepened our appreciation for an aspect of punishment, as we understand it, which transcends its prosaic role as a mere law-enforcement mechanism. As we saw, Kant set out to articulate what we really mean

when we say that criminals should pay their “debt” to society, and when we say that society ought to treat criminals as responsible moral beings rather than as erring children.

Kant paints a picture of the moral life as guided by a hope for perfect justice as it would be dispensed by an avenging God. This account of the human longing for justice implies that what retributive punishment fundamentally means to us is not a restriction on deterrence, or satisfaction of the victim’s honor, or even the expression of society’s notions of right and wrong. Above all, to punish is to bring into worldly existence the dependence of happiness on virtue. The longing to see the virtuous flourish and the wicked punished shows its power by enduring in the soul despite Hobbes’s systematic attempt to uproot it, and by giving rise to the belief in a punishing God (as Kant candidly admits). By enduring in the soul, the retributive outlook defies modern political philosophy’s attempts to tame religious longings by the promise of happiness in this life. We owe Kant a debt of gratitude for his acute sensitivity to the human love of justice and for his unflinching willingness to give it theoretical articulation. But can we remain satisfied with Kant’s view despite the inadequacies we have uncovered in it?

I have attempted to show that Kant’s defense of the morality of retribution from the point of view of modern political rationalism fails, and that in order to fully understand it one must appreciate the influence on Kant of the classical tradition. Does this mean that the aim—pursued by Hobbes, Kant, and by all of their contemporary successors—of elaborating a moral theory of punishment on the basis of modern principles is unattainable? Must we choose between accepting Hobbes’s callous view of punishment and abandoning rationalism altogether for a more morally satisfying—but

also more mysterious—Biblical faith? In the next two chapters I will suggest an alternative solution. Instead of abandoning the rationalist approach to the question of punishment altogether, we may still investigate whether a moral theory of punishment can be grounded on the principles of classical political rationalism.

Chapter 4: The Principles of Penal Reform in Plato's *Laws* V

The development of modern penal thought from Hobbes to Kant shows the importance of giving moral responsibility, moral judgment, and desert their due. Hobbes had attempted to purge these elements from modern penal practice altogether in accordance with his new understanding of crime, but he found that he could not do so without equivocation. Kant later tried to recover these elements in his theory of punishment and to reestablish them on modern principles, but as that task appeared to him to be difficult, if not altogether impossible, he was compelled to draw on the insights of a premodern view of crime and punishment—namely, that of the Bible. Kant thus failed to accomplish his goal of grounding a moral theory of punishment on *modern* principles. I believe it would not be misleading or unfair to say that, with very few exceptions, contemporary Western penal theorists continue to conduct the punishment debate within the range of theoretical possibilities first defined by Hobbes and Kant. It is not surprising, then, that they too have been unable to provide a remedy for the moral crisis in modern penal justice. In light of all this, it is not unreasonable to suggest that we look outside of the modern tradition for other possible solutions.

There appears to be good reason to turn to the classical tradition in particular. Kant had already turned to classical thought (in its Biblical expression) for guidance when he found the modern theoretical framework not entirely hospitable to retribution, and yet he wanted to make use of classical insights without abandoning his commitment to modern principles, which, as I have argued, was a large reason for his failure. But

perhaps we can succeed where Kant failed if we try to follow his instincts in a more thoroughgoing way, that is, by taking the classical approach to the question of punishment from the beginning.

One of the most profound treatments of the topic of punishment in all of classical philosophy can be found in Book IX of Plato's *Laws*. The *Laws* is unique among Plato's dialogues in that it is his most political work. Plato's other great dialogue about politics, the *Republic*, is a dramatic portrayal of the philosopher Socrates constructing a city in speech in the company of men the majority of whom would not normally be regarded as public-spirited. In such circumstances, Socrates is permitted to be unreserved in his reflections about politics. He therefore proceeds not to teach statesmanly prudence but to show how politics appears to a philosopher like himself, and to indicate the substantial reasons why he never willingly engaged in political activity (cf. 514a-519c with *Apology* 31c-32e). The *Laws*, on the other hand, portrays another philosopher—the nameless Athenian “stranger”—in a solemn setting elaborating a legal code on behalf of a founder-to-be of an actual city, and thus taking on an emphatically political role. Unlike Socrates, the Stranger proceeds in such a way as to never allow his philosophical attitude to take center stage and dominate the course of the discussion or its results. For the sake of the conversation the Stranger takes for granted the ends politics sets for itself—freedom, self-sufficiency, civic virtue, and the like—and almost never questions them openly (but cf. 804a-c).¹ In accordance with this procedure, the Stranger accepts the ends of political life

¹ I have relied on the Loeb edition of the *Laws*, edited by Jeffrey Henderson. Some of the shorter translations are my own, but for longer translations I have relied heavily on that of Thomas Pangle, *The Laws of Plato*.

as they appear to political men rather than as they would appear to a philosopher—or to anyone else who is not committed to political life. These ends are given expression by the Stranger's two interlocutors, the aged Dorian gentlemen Kleinias and Megillus, both of whom come from cities where ancestral laws are revered and where they hold important positions of leadership and responsibility. In fact, Kleinias is revealed in the course of the discussion to have been appointed by his city of Knossos as one of a committee of ten legislators to draw up a legal code for a new Greek colony, and it is on his behalf that the Stranger elaborates his system of laws (702b-702e).

In this, the Stranger's approach to politics in general and, as we shall see, to punishment in particular, stands in contrast to the approach taken by Hobbes and Kant. Both Hobbes and Kant gave simplified accounts of the complex civic view of crime and penal justice and tried to deduce their respective theories of punishment from a single theoretical principle—the right to self-preservation in Hobbes, the categorical imperative in Kant. Conversely, at the urging of his interlocutors the Stranger accepts (1) that the political community has a right to secure itself against crime, (2) that justice must always benefit and not harm those to whom it is done, and (3) that criminals should be punished as they deserve—i.e., retributively. The Stranger accepts all of these needs and demands of the political community despite the fact that they can be shown upon closer examination to be in tension with one another, both in practice and in theory. The Stranger elaborates his penal code in such a way as to comprehend and reconcile these occasionally divergent concerns as much as possible, and as consistently as possible, thereby ensuring that his penal code will be acceptable to ordinary citizens in the highest

degree. The Stranger thereby accomplishes what modern penal thought has since failed to accomplish: he elaborates a legal framework or set of rules for a penal system that could be both effective and morally satisfying.

It is easy to see, however, that this approach has its own limitations. It goes without saying that any attempt to compromise with our moral principles for the sake of political convenience should be regarded with suspicion, since it threatens to undermine our effort to work out a rational and humane theory of punishment in light of the truth about crime and justice. In these final two chapters, I will argue that one of the major reasons why the Stranger takes this approach is that he believes a perfectly satisfying normative theory of punishment—one that is both entirely noble and practicable—to be impossible. Nevertheless, two important qualifications must immediately be made. First, to say that a perfectly humane normative theory of punishment is impossible does not yet mean that all penal codes are morally equivalent. Some penal codes can be more humane than others, and I will try to make the case that the Stranger's penal code is a good example of the former. As such, it may be used as a pattern (appropriately adapted to particular circumstances) by thoughtful judges, legislators, and statesmen. (For example, the understanding of the nature of crime that informs the Stranger's account of penal laws could well be canvassed by the U.S. Supreme Court in framing its opinions in Fifth, Sixth, and Eighth Amendment cases.) Second, the Stranger does not merely assume that a perfectly humane and rational penal code is impossible but shows *why* it is impossible. In subtle but clear ways, the Stranger draws our attention to crucial reflections, which he does not always make explicit and thus requires readers to arrive at on their own, about

certain permanent features of political life which force him to introduce inconsistencies into his penal code. Through the dramatic device of the Stranger's dialogue about penal laws Plato both teaches his readers about the nature and limits of political reform and shows how a penal system might be organized in practice to deal with crime as sensibly and humanely as possible, and without utopian hopes.

The Athenian Stranger's Approach to Legislation and the Two Penal Preludes

The Stranger tries to arrange a legal code by means of which the best practicable city could come into being. A city can correctly be called best only if it is both stable (self-sufficient, strong, and united) as well as committed to promoting the virtue of its citizens. The virtue of citizens must not be understood as subordinate to the stability of the city, but as an end in itself. Rather, the reverse is true: war and domestic order must be regarded as being in the service of peace and the exercise of virtue as a whole. Thus the martial virtues and the virtue of loyalty to one's regime must be regarded as subordinate to the highest virtue, prudence or intelligence (cf. 628c-e with 631b-d). Of course, it is not likely that all or even most citizens will be able to attain the whole of virtue; at best, only a few will be able to attain it (cf. 627c). The difference in virtue among the citizens is a potential source of conflict and civil strife, as well as of vulnerability to external enemies, things the legislator must try to forestall by instilling friendship in the city between the more virtuous and the less virtuous citizens (cf. 627c-628c with 690a-d). Such friendship is secured primarily by establishing the rule of law in

place of the rule of men, even if cities ruled by law are only second best with a view to virtue (cf. 627e-628a with 739c-e).²

According to the Stranger, if the future city is to be as stable and also as accommodating to human virtue as possible then the laws must be accompanied by preambles or preludes (*prooimia*) which communicate to citizens the wise purpose of the laws and encourage them to obey willingly.³ The preludes contain both rational arguments and exhortations. Thus, in the first place, the preludes attempt to educate citizens as to the nature of virtue as a whole and to show how the laws are designed to promote it, but insofar as rational education fails, the laws are also able to promote obedience through exhortation and admonition. To the extent that it is successful, this twofold role of the preludes allows for the effective rule of law and, to some degree, of intelligence as the principle embodied in law. The result of the Stranger's artful weaving together of argument and exhortation is to make the preludes highly complex in their structure. This complexity is intended to provoke thought and to instruct citizens, to the extent that they are capable of following the subtle indications and clues left by the legislator, about the highest subjects of legislation. The preludes are at times even contradictory, as their two distinct functions—rational education and exhortation—occasionally make the law appear to intend different things.

By examining carefully the preludes to the Stranger's penal laws we can hope to gain insight into his understanding of the purpose of punishment. There are two sections of the dialogue that discuss the principles of punishment. In Books IV and V, the

² Cf. Leo Strauss, *The Argument and the Action of Plato's 'Laws,'* 5.

³ The present paragraph is a summary of the section extending from 718b to 723d.

Stranger delivers a long speech, running from 715e to 734e, which serves as a general prelude to the laws, and in which punishment is one of the topics taken up (728b-d, 730d-732b). The principles of punishment are again discussed in Book IX just before the Stranger elaborates his penal code. The long prelude and a few short sections of the discussion of penal law in Book IX (854b-c, 870a-871a), are explicitly addressed to the citizens of the new colony. What the Athenian says in these passages, then, is his most public justification of his penal laws. On the other hand, most of the discussion of punishment occurring in Book IX is carried on privately among the three interlocutors and is not said to be intended for a public audience. Yet this private discussion, like the whole conversation conducted in the *Laws*, is still a kind of prelude, since at one point the Athenian states that the highest magistrates of the new city will have access to a (presumably transcribed) version of the whole dialogue that takes place in the *Laws* as a “model” (*paradeigma*) to be studied (811c-e).

The Public Penal Prelude and Divine Punishment

The long prelude is divided into two sections. The first section (715e-718a) deals with the proper way of honoring gods, heroes, and ancestors. The second section—containing the discussion of punishment—addresses citizens regarding the right treatment of one’s soul, body, and property (726a-734e). Punishment, then, concerns what is one’s own. Among the things that are one’s own, the highest and most divine is the soul. The goodness of the soul consists not in having its desires fulfilled indiscriminately, but in practicing those things which the lawgiver praises as noble and abstaining from things

which he blames as base (727b, 727c-d). Lower in rank than the soul are the goods of the body, including life itself, and wealth (727d-728a). Those who consider survival to be always good reckon incorrectly because they assume that “everything done in Hades is bad,” when in fact “one doesn’t know whether things aren’t just the opposite—it may be that among the gods below are to be found the things that are by nature the greatest goods of all for us” (727d). It is thereby implied that insofar as the lawgiver praises as noble the risking of one’s life in defense of the city this ought to be considered preferable to, and more in accord with one’s own good than, fleeing battle in order to preserve oneself. Moreover, since, as the prelude asserts, there are gods in Hades citizens are implicitly advised to fear disobeying the legislator and incurring a terrible punishment in the afterlife (cf. 716a-b). One might even suspect that this threat of divine punishment after death would be more likely to induce conformity to the laws than the legislator’s praise of noble actions.

We can see that the prelude merely asserts without discussion (1) that the goods of the soul are higher than those of the body, and higher even than life itself, (2) that the goodness of the soul consists in following the advice of the lawgiver, (3) and that there is a subterranean life after death in Hades which may be better for men than the goods enjoyed in this life. In this way, the Stranger prepares a much more conventionally Greek civic outlook than the one Socrates attempts to instill in the guardian class in the *Republic*. The Stranger’s prelude engenders a pious reverence for law rooted in the belief in life after death and divine punishment, whereas Socrates not only purged from his city in speech those eschatological elements of conventional Greek religion which made the

afterlife frightening, but even refrained from saying whether or not there was an afterlife at all (*Republic* 386a ff.). One might look at the *Republic* as Plato's attempt to imagine a political community in which the fear of divine punishment was not needed to make men just.⁴ The failure of that attempt in the *Republic*, and the return to conventional piety in the *Laws*, suggests that Plato did not believe the policy of eradicating the fear of divine punishment to be politically advisable. We might say that Plato agreed with Kant in at least this one respect: that when men act morally they cannot help but affirm the existence of an avenging deity.

The prelude chastises citizens for presuming to know whether the things in Hades are good or bad for them. But if the citizens do not know one way or the other about the things that are in Hades, then how would the human legislator—i.e., the Stranger—know any better? On the one hand, we would certainly expect a general presumption of the legislator's superior wisdom on the part of the citizenry. The conversation in the *Laws* is elaborated on the way to the cave and temple of Zeus near Knossos, and upon their return the three interlocutors might give the impression that the Stranger's legal code was a product of divine inspiration (cf. 625a-b). And yet Plato's readers (as well as the highest magistrates of the city who will have access to the *Laws*) would also know that the Stranger's legal code was conceived *before* the three men arrived at their destination, and so the effect of the open-endedness of the prelude's suggestion that "one does not know" about the things in Hades is in no way diminished. In this very inconspicuous way, and

⁴ Cf. Allan Bloom, *The Republic of Plato*, "Interpretive Essay," 353-355.

without greatly disturbing the city's presumption of the legislator's wisdom, the prelude calls into question the existence of divine punishment in the afterlife.

Punishment and the Goodness of Justice

After its statements regarding the afterlife, the prelude shifts to the topic of crime and civil punishment. Since the laws guide citizens to virtue, which is their own greatest good, it follows that those who disobey the laws are harming themselves. Accordingly, the prelude states that “the judicial penalty [*dikē*] for wrongdoing that is said to be the greatest is to become similar to men who are wicked, and, in becoming similar, to avoid good men and be cut off from good conversation, and instead to attach oneself to the bad by seeking intercourse with them” (728b). The greatest punishment for injustice, then, is not the harsh treatment of criminals by the city or by the gods, as is commonly thought, but the effect that injustice has on the soul of the criminal and on the character of his relations with his fellows. The first word of the penal prelude on punishment, then, is that citizens should willingly seek to be just not in order to avoid the legally prescribed penalty, but because the just life is better than the unjust life.

In order for this to be true, however, justice would first have to be *proved* to be better for the just man. The requisite proof for this important premise is apparently given a page later in the long prelude, where the life of the virtues is compared with the life of the vices with respect to their relative amounts of pleasure and pain. According to the prelude, we are by nature capable of wishing only for a kind of life in which on balance pains are exceeded by pleasures (733b-d). It is on this basis that the virtuous life is said to

be superior, as it is said to be not only nobler than the vicious life but also more pleasant, at least “if one tastes it in the correct way” (733a). Specifically, the prelude concludes that the moderate, prudent, and courageous lives are more pleasant than their opposites. Yet the prelude makes no mention of justice, thus leaving it as an open question whether or not justice is more pleasant than injustice (733d-734e). The way in which the *Laws*, as distinguished from the *Republic*, solves this problem is by defining justice as a mixture of the other three cardinal virtues—prudence, moderation, and courage—rather than as an independent virtue in its own right (cf. 631c-d with *Republic* 433a-b and 443c-e). By assimilating justice to a mixture of the other virtues the *Laws* abstracts from the self-sacrificial aspect of justice which in the *Republic* is expressed as doing one’s own job or duty (i.e., for its own sake).

To return to the passage on punishment in the long prelude which we have been discussing, it is now stated that the harmful effects of injustice for the unjust man (assimilation to the bad and exclusion from association with the good) are not just punishment but “retribution” (*timōria*), “since what is just [*to dikaion*]—including judicial punishment [*dike*]—is noble” (728c). Unlike justice, which is noble and choiceworthy, retribution “is a consequence [*akolouthos*] of injustice” that makes the man not better but worse. “Both he who undergoes this [i.e., retribution],” the prelude continues, “and he who escapes it are miserable: one because he does not get cured, the other because he is destroyed in order that many others may be saved” (ibid.). With this, the prelude defines justice according to a very high standard of benevolence, excluding not only retribution but even certain cases of deterrence from the category of just

punishments. The specific consequence of injustice which the prelude calls “retribution” cannot be considered just for the simple reason that it leaves the wicked man to suffer a miserable life of injustice. Moreover, even when the criminal escapes retribution by being punished to deter others, such punishment cannot be considered as just in the strict sense, regardless of how much the rest of the city might stand to benefit from it, because the one who is destroyed is not benefited.⁵ When the city punishes an offender to deter others it is clear that it does not *wish* to do so for its own sake but does it only out of a need to secure itself against further crime; it does not see deterrence itself as noble. In this way, the penal prelude acknowledges the admirable tendency of human beings to feel aversion toward sacrificing individuals to the collective interest of the rest of society. According to this standard, deterrent punishment would only be just when the injustice of the criminal is incurable, and when it has so corrupted his soul that killing or otherwise incapacitating him would be a lesser evil for him than continuing to live an unjust life. In all other cases, the only truly just punishment would be the one that somehow rehabilitates the criminal (cf. *Apology* 25c-26a; *Lovers* 137b-d).

The prelude does not, however, say that deterrence would be *unjust* if it failed to benefit the individual who is punished as an example to others—it merely denies that it is just. The prelude thus seems to leave room for understanding deterrent punishment as a necessary concession to the exigency of crime. But by refusing to call such punishment

⁵ There is some controversy over the correct interpretation of the passage running from 728b2-c5. I follow E. B. England’s reading of *timōria* as referring only to the criminal’s acquisition of a bad character and his falling into bad company, and reading “he who escapes it” as referring to the man who is killed by his city as an example to others, since by being destroyed he is released from the misery of having a bad character. See *Laws of Plato*, vol. 2, 476-478. Mackenzie, on the other hand, reads *timōria* here as referring to both the capital punishment imposed by the city and the wretched condition of the soul of the criminal as a consequence of his assimilation to bad men. See Mary Mackenzie, *Plato on Punishment*, 196 n. 62.

just the prelude also tries to moderate human beings' naturally strong concern for the mere security of their persons and property and to encourage them to care more for the souls of those of their fellow citizens who have fallen into crime.

But does the city only punish for the sake of deterrence, and thus reluctantly, or doesn't it also or at the same time get angry with criminals, punishing them willingly and even enthusiastically? The prelude's claim that retribution is not just or noble because it harms criminals instead of benefiting them is rather odd, since retributive punishment is at least sometimes thought to be just and even noble precisely *because* it harms criminals and thereby treats them as they justly deserve: in subjecting the criminal to a fitting amount of suffering retributive punishment is thought to vindicate justice itself. Equally odd is the prelude's assertion that retribution is a natural consequence of injustice rather than an intentional one. But is this not plainly false? While it may be true that the corruption of one's character is inseparable from the habitual performance of unjust actions, and thus unavoidable, it is far from obvious that the social consequences of injustice—the exclusion of the criminal from good society and his destruction—are natural. In other words, is it not obvious that retribution is an entirely human, and entirely intentional, phenomenon? Much like the English word “retribution,” the Greek word *timōria* has a strong connotation of intentional and deliberate harming of an offender in view of his guilt. Why then does the Stranger here simply ignore the whole question of why human beings choose to punish retributively? The reason for this becomes clearer in the conclusion of the public penal prelude.

“Every Real Man Must Be of the Spirited Type”

The prelude’s monologue on punishment next turns to deal with the issue of spirited punishment. It now ranks citizens according to their dispositions towards crime. He who refrains from doing injustice is honorable, but he who works to prevent unjust men from doing injustice by informing the magistrates of the injustice of others is more honorable still. “Yet the great man in the city, the man who is to be proclaimed perfect and the bearer of victory in virtue, is the one who does what he can to assist the magistrates in inflicting punishment [*sunkolazon*]” (730d). Shortly thereafter, we are surprised to hear that these men who are highly praised by the prelude for their virtuous assistance to the magistrates in punishing crime will be punishing *retributively*, for the prelude goes on to declare that “every real man should be of the spirited type [*thumoeide*]” (731b)—the type of man who has a tendency of being moved by anger and indignation (cf. 731d6). On the basis of the prelude’s earlier definition of just punishment, we would have expected the prelude to say that a virtuous man would try as much as possible to rehabilitate criminals, and when this is impossible to try to make examples of only those who are incurable, but certainly never to punish retributively. What, then, explains the prelude’s departure from its own standard?

The passion of spiritedness (*thumos*) plays a very important political role in Plato’s dialogues. Its central character—the concern with one’s own—lies at the heart of a number of distinctively human emotions and mental states, some noble and some base,

that include love of honor, civic and familial loyalty, courage, shame, and indignation.⁶ Spiritedness is *the* political passion because without it political life would not be possible: one cannot have a city without loyal citizens, brave soldiers, and attentive parents. But despite, or rather because of, the intense concern with one's own which it fuels, spiritedness can too narrowly focus our view on what is our own—especially our honor or dignity—so much so as to blind us to reason and to what is truly good for us, for our friends, and for our city (cf. 731e-732b).⁷ The self-assertive tendency of spiritedness which causes us to demand the respect we are owed and to lose sight of the good leads us to want to harm those who have dishonored us, even if this would not advance our true interests in any way. From the spirited perspective, every unlawful or undeserved injury is viewed as a loss of honor for the victim and as a “gain” of honor for the offender.⁸ Spirited indignation is guided by the thought that it can restore this loss of honor by inflicting an equivalent injury, and thus dishonor, on the original offender. It is no matter, from the point of view of spirited indignation, that the future interests of the two parties are lost sight of in the course of this retributive settling of accounts. By indicating that the man declared by the city to be virtuous will be punishing out of spirited anger, the prelude shows its awareness of the fact that its authoritative affirmation of the absolute benevolence of justice will not be able to transform the essentially spirited character of the citizenry. It thereby acknowledges the need to tolerate to some extent the retributive

⁶ Cf. Thomas Pangle, “The Political Psychology of Religion in Plato’s *Laws*,” 1062-64. See also Aristotle, *Politics* 1327b38-1328a17.

⁷ Rousseau’s description of *amour-propre* is in all essentials in agreement with Plato’s understanding of *thumos*. See Rousseau, *Judge of Jean-Jacques: Dialogues*, 9.

⁸ Cf. Aristotle, *Nicomachean Ethics* 1132a9-15.

manifestation of spiritedness, even while it tries to temper this manifestation as much as possible.

The prelude proceeds to explain that every real man should be of the spirited type because “there is no way to avoid those injustices done by others that are both dangerous and difficult, or even impossible, to cure, except to fight and defend oneself victoriously, in no way easing up on punishment[;] this every soul is unable to do, if it lacks a high-born spiritedness” (731b). Effective law-enforcement in a city where the citizens are their own police force requires the vigilance and active help of all. But this is difficult to accomplish because of what we might call the ‘collective action problem’: since active participation in law-enforcement is not only time consuming but also potentially dangerous, few citizens will want to get personally involved and would rather leave the job to others. This is why the legislator must encourage a zeal for punishment among the citizens and does so by praising it as a virtuous activity. But while the prelude appears to praise the punitive citizen, it is careful not to call him “just,” nor to designate as “justice” the punishment that he metes out.⁹ Moreover, what the Stranger actually says is not that the punitive man *is* virtuous, but that he is “*to be proclaimed* perfect and the bearer of victory in virtue.” In light of the prelude’s definition of just punishment, carrying out spirited punishment with the intention of doing harm cannot be considered virtuous in the strict sense.

⁹ From 730d1 the Athenian never uses *dike*, but only *kolasis* (or a cognate) to designate punishment (730d8, 731b7). This word can mean to punish, chastise, or correct, and does not carry with it the sense of being necessarily in accordance with right.

In order to temper the punitive civic spirit still further, the prelude goes on to say that every real man must be of the spirited type “yet also as gentle as possible.”¹⁰ “In regard to the curable injustices men commit,” the prelude explains,

one must first understand that no unjust man is ever voluntarily unjust. For no one anywhere would ever voluntarily acquire any of the greatest evils—least of all when the evil afflicts his most honored possessions. Now the soul, as we asserted, is truly the most honorable thing for everyone; therefore no one would ever voluntarily take the greatest evil into his most honorable possession and keep it for the rest of his life. So the unjust man, like the man who possesses bad things, is pitiable in every way, and it is permissible to pity such a man when his illness is curable; in this case one can become gentle, by restraining one’s spiritedness and not keeping up that bitter, woman’s raging. But against the purely evil, perverted man who cannot be corrected, one must let one’s anger have free rein. This is why we declare that it is fitting for the good man to be of the spirited type and also gentle, as each occasion arises. (731c-d)

In this exhortation to gentleness, the prelude once again relies on the principle established earlier—that injustice harms the soul of the unjust man—and on this basis concludes that injustice must be involuntary when it is committed by unjust men who are curable. If justice is the good condition of the soul, and if injustice deprives us of this good, then no one who knows what he is doing would ever choose to be unjust. If someone commits

¹⁰ For an extensive discussion of the various ways in which the Stranger’s institutions in the *Laws* are designed to temper spirited anger see Lorraine Pangle, “Moral and Criminal Responsibility in Plato’s *Laws*.”

injustice thinking that he will profit thereby, we must conclude that he is mistaken, based on what the prelude says, because he does not understand that the harm to his soul that will result will far outweigh any profit he might gain through crime. The choice made in such a case would therefore not be a free one in the full sense, if a free choice presupposes full knowledge of what one is doing. In view of this, the prelude goes on to say, as we just saw, that curable offenders deserve pity and gentle correction, as do those who go astray and unintentionally harm themselves, rather than angry retribution.

At the same time, the prelude clearly implies that incurable criminals—those who are thoroughly evil and cannot be corrected—do act voluntarily, since it is said that against such men it is permissible to give one's anger free rein. It is unclear why the prelude draws a distinction between curable and incurable criminals and identifies only the former class as acting involuntarily. The prelude does not explain why exactly the argument about the badness of injustice does not apply to the second class as well. Does the prelude mean to say that the curable criminal is the one who believes that justice is good and choiceworthy for him and who regrets his momentary lapse in judgment, whereas the incurable criminal is the one who has contempt for justice and so does injustice without repentance? But does this establish that the unrepentant man acts voluntarily? The prelude maintains that injustice is the greatest evil that can afflict the soul, and if this is true, then the incurably unjust man would seem to be gravely mistaken in his contempt for justice, not knowing what is truly good for him, and if he doesn't know this one thing of utmost importance to him, then did he not act in ignorance and therefore involuntarily? Or does the prelude mean that the incurable criminal does what

he knows to be bad for him? We cannot answer these questions with any confidence until we have a more thorough account of the nature of criminality than the one given in the public penal prelude. Such an account is provided in Book IX of the *Laws*, to which we will turn in the next chapter.

Chapter 5: Plato's Philosophical Criminology in *Laws* IX

In the extensive section of the *Laws* following the long prelude and leading up to the end of Book VIII, the Stranger devotes himself to the task of legislating for all aspects of civil life in the future Greek colony. For each subject of legislation—including the election of public officials, education, festivals, and farming—the Stranger indicated which actions would have to be prohibited and suggested the kinds of penalties that would be appropriate for transgressions of those prohibitions. In Books IX and X, the Stranger brings his penal legislation to a conclusion by establishing penal laws for the most serious kinds of crime—temple robbery, sedition, treason, murder, assault, theft, and impiety.

Book IX begins with the Stranger announcing that “some of the matters that require judicial punishments (*dikas*) have been discussed—those that pertain to farming and whichever follow them—but with regard to the greatest matters there has been as yet no discussion, showing in each single case what retribution (*timōriai*) should be attached . . . These are the things that should be stated next after those others” (853a). The way in which the Stranger formulates his proposal provokes an obvious question. Why does he refer here to *timōriai* as a subset of the *dikai* when the public penal prelude had clearly excluded retribution from the class of the *dikai*, since, unlike justice, retribution does the criminal harm?

To make sense of this inconsistent use of the word *timōria* we must recall the kind of civic outlook on crime that had been encouraged earlier in the *Laws* by the public penal prelude. As we saw in the last chapter, the earlier prelude called attention to the

need for all citizens to help the magistrates enforce the laws, and to do so with zeal. While it admonished citizens to take pity on the curable criminals, it also urged them to give free rein to their spirited anger against the incurable ones, and even went so far as to crown the most vigilant, punitive citizens as “real men” and best with respect to virtue. The political necessity of providing for effective law-enforcement in a small republican community compelled the prelude to tolerate spirited anger, despite the fact that such anger tends to lead human beings to punish retributively, which was said to be contrary to the prelude’s fundamental stance that justice must benefit and not harm. In blurring the difference between *timōria* and *dikē* at the opening of the discussion in Book IX, the Stranger echoes the confused but politically salutary opinions of the spirited citizens praised in the earlier penal prelude. While the Stranger thus continues in Book IX with his strategy of accommodating rather than openly questioning the spirited attitude of the city towards crime, he will also attempt, as we shall see, to conduct a parallel discussion of criminality on a much more critical level. In this way, the Stranger tries to address the fundamental question of the nature of criminality that had been left unanswered by the public penal prelude.

In his next remarks, the Stranger says that it is in a certain sense shameful even to institute punishment in a city that is well administered and correctly equipped for the practice of virtue (853b-c). The regime described in the *Laws* is one committed to the cultivation of virtue to an unprecedented degree. Not just its educational system but almost all of its institutions are designed to promote the acquisition and life-long practice of the virtues of law-abidingness, self-restraint, and piety. The Stranger’s regime may be

more committed to the cultivation of civic virtue than any political community that has ever existed. In such a city, the Stranger suggests, crime should be an anomaly rather than a common occurrence, and therefore penal laws should be unnecessary. The fact that the Stranger now finds himself compelled to provide a penal code is shameful for him as a legislator because it implies a recognition on his part that his legislation will fail to accomplish its major task. Fortunately, there is a circumstance that excuses the Stranger's decision to lay down penal laws: the weakness of human nature. Human nature is responsible for making some men so resistant to education that even the strongest laws cannot "melt" them (853c-d; cf. 854a). Unlike himself and Kleinias, who are merely human legislators legislating for human beings, the ancient lawgivers who gave laws to heroes (the children of gods) were not saddled with the shameful task of framing penal laws, apparently because the divine natures of the heroes were so good that they were perfectly receptive to a legally prescribed education in virtue which made it unnecessary to compel them to be just through punishment (cf. 853c). Not being in the fortunate position of the ancient legislators, the Stranger will have no choice but to supplement civic education with punishment.

The Stranger assumes here that crime is always the result of a poor or stubborn nature that resists the laws' educational effect, and he points to the justice of the good-natured heroes as his evidence. But what he says disagrees in fact with the best known and most authoritative reports among the Greeks about the deeds of gods and heroes—such as the revered poems of Hesiod and Homer. Is it likely that the Stranger has temporarily forgotten about the many crimes that the gods themselves are reported to

have committed against one another, to say nothing of the heroes (cf. *Euthyphro* 5d-6a)? Contrary to what the Stranger says, the authoritative reports teach that even the best and most gifted natures—i.e., the divine natures of gods and heroes—can do injustice when they believe they have good reason for it. These reasons can be either noble, as in the case of Prometheus’ benevolence, or base, as in the case of Zeus’ lusts. If we cannot reasonably suppose that the Stranger has forgotten about the crimes of the gods and heroes, then we must interpret his paradoxical claim about criminality here as a puzzle or provocation. By presenting this puzzle at the beginning of the discussion of Book IX, the Stranger focuses our attention on the central question of the true causes of crime. Having tacitly pointed to this question, the Stranger proceeds to the topic of capital crime.

Capital Crimes

The three capital crimes identified by the Stranger are temple robbery (a most serious crime for the Greeks), the overturning of the laws, and treason. All three are to be tried by the same judicial procedure (described by the Stranger at 855c-856a) and punished in the same way (856e-857a). If a domestic servant or a resident alien is caught committing a capital crime he is to be whipped as much as the judges should decide is appropriate and thrown naked beyond the borders of the country, but not killed, since it is possible that by undergoing punishment he will become better, “for no judicial punishment that takes place according to law aims at what is bad” (854d). In this way, the Stranger reaffirms the principle he had established in the public penal prelude that justice is always directed towards the good. The Stranger also continues speaking here of the

cause of crime in the soul as a “disease” that may or may not be curable (cf. again 731c-d). If it is curable, then the criminal can hope to be cured by means of a chastising punishment, if not, then execution or suicide is preferable to his continuing to live with a diseased soul (854b-c). After the Stranger completes his discussion of the correct judicial procedure and punishment for domestic servants and resident aliens, he moves on to discuss the judicial procedure for full citizens. A full citizen convicted of a capital crime should be punished with death, for “the judge should think of this man as already incurable” since the rigorous civic education he received was not enough to prevent him from committing the greatest crime (854e). The penalty of death should be regarded as a release from his miserable condition and as the least of evils for him. From this point of view, the Stranger can still speak of capital punishment as a *dikē*, as something that aims at what is good.

Yet in spite of this, the Stranger continues in this section to refer to punishment as *timōria* and *dikē* interchangeably, as he had at the very beginning of the discussion and in Book V. In order to encourage magistrates to do their duty in punishing criminals, the Stranger admonishes that if a magistrate “holding one of the highest offices in the city lets these things escape his notice—or, not because they escape his notice, but because of cowardice, fails to wreak retribution (*timōria*) on behalf of his own fatherland—such a citizen must be held to be second in evil” (856b-c; cf. d). While we know the reason why the Stranger persists in tolerating this major inconsistency in the civic understanding of punishment, we do not yet know what the Stranger’s two interlocutors think about this inconsistency, as neither has voiced any opinion on it one way or the other either here or

in the context of the public penal prelude of Book V. In fact, we do not yet know whether Kleinias and Megillus have noticed the tension at all.

Both men are experienced in public life, and, at least in Kleinias' case, experience has taught him to doubt at least some of the authoritative beliefs openly professed in public (cf. 624a-b, 625c-626d). It is therefore possible that when Kleinias had acquiesced in the Stranger's policy of accommodating punitive spiritedness in the city, first in the public prelude and now here in Book IX, he had done so with an eye to its political utility—that is, for the same reasons that were advanced by the Stranger himself. It is also possible, however, that Kleinias had acquiesced in this policy for an entirely different reason: because he shares the spirited civic attitude towards crime. In order to find out how things actually stand with Kleinias, the Stranger decides to provoke him into revealing his genuine views about retribution—if only to confirm for himself whether or not Kleinias has indeed grasped his policy regarding punishment. Yet, as we shall soon see, in bringing to light Kleinias' real views on retribution the Stranger has another goal in mind as well—to steer their conversation in the direction of a theoretical discussion of the nature of crime.

The Punishment of Theft and Kleinias' Objection

In contrast to the lengthy treatment of capital crime, the Stranger's dispensation regarding theft in general is remarkably terse. "As to stealing, whether someone steals something great or something small," the Stranger says, "let there be one law and one judicial retribution [*dikēs timōria*] in all cases" (857a): double the value of the stolen item

is to be paid as a fine; if the convicted offender does not possess enough property to pay the fine, he is to be imprisoned until he does pay or until he persuades the successful prosecutor (857a-b). In response, Kleinias earnestly objects to the Stranger's proposal:

How comes it, stranger, that we're saying it makes no difference to the thief whether he's convicted of stealing something great or something small, whether from sacred or profane places [*ex hieron ē hosion*], or whatever other respects in which a theft can be entirely dissimilar? For crimes that are thus various, is the lawgiver in no way to follow with penalties of a similar variety? (857b)

Kleinias' objection is provoked by the Stranger's attempt to make the penalties for theft strictly proportional to the material value of the stolen goods, regardless of what other significant qualities those goods may possess. By putting special emphasis on the difference between the sacred and the profane, Kleinias indicates that he thinks the significance of at least some types of theft is not reducible to the mere loss of wealth. As we saw from the Stranger's discussion of capital crimes, temple robbery was regarded as a very great injustice, and although the Greeks distinguished between temple robbery and stealing sacred funds from places outside the temple,¹ we can safely assume that the seriousness of committing an injustice against both men and gods is common to both crimes. Kleinias seems to want thieves guilty of both injustice and impiety to be punished

¹ Cf. Pangle, *The Laws of Plato*, 533, n.7.

in proportion to the greater moral wickedness of their crime, rather than in proportion to its merely pecuniary aspect.

It is still possible that Kleinias is simply being the foresighted statesman and reminding the Stranger of the city's need to impose harsher punishments for theft of sacred property in order to discourage impiety, since impiety might be expected to have a generally subversive effect on the citizens' respect for their laws. Or is there something more to Kleinias' objection? The other possibility is that Kleinias' sudden interjection in defense of more severe punishments for theft of sacred property is rooted in his concern with just desert for its own sake. In other words, it is possible that Kleinias' reaction is a manifestation of a genuine concern with retributive punishment. Moreover, if Kleinias' emergent concern is indeed with retribution, it is unclear how it would be connected with the difference between the sacred and profane—i.e., how it would be related to piety and religion. In order to get at what exactly is behind Kleinias' objection, the Stranger temporarily puts aside the task of legislating and guides the conversation into more critical territory.

The Stranger next praises Kleinias for his objection and claims that it has somehow reminded him that “what pertains to the laying down of laws has never been worked out correctly in any way” (857c). Recalling the earlier discussion about the need for laws to be accompanied by preludes, the Stranger now says that human beings living under all existing legislation resemble slaves being doctored by slaves (cf. 719c-723d). Unlike slave doctors, who practice medicine on the basis of experience rather than knowledge and who impose treatments on their patients without persuasion or

explanation, free doctors practice their art by carrying on a dialogue with their free patients using arguments “that come close to philosophizing, grasping the disease from its source, and going back up to the whole nature of bodies” (857c-d). This is the first instance of the word “philosophy” or one of its cognates in the whole work, which further marks the dialogue’s ascent from the more mundane subjects of legislation to a theoretical plane.² The Stranger now reflects that if one of the slave doctors should come across the free doctor with his patient, he would burst out laughing and declare, “Idiot! You’re not doctoring the sick man, you’re practically educating him, as if what he needed were to become a doctor, rather than healthy!” (857d-e). The Stranger’s approval of the educational approach of the free doctor in the context of a discussion of penal laws seems to be intended to recommend a similarly educational or philosophic approach to crime. If injustice is analogous to disease, then the proper response to crime—i.e., what the criminal deserves—would be some kind of healing. And if the healing most appropriate to a free patient requires the doctor to procure the patient’s informed consent by educating him about the nature of the disease and the proposed cure, then by analogy the appropriate response to crime would be to educate the criminal about the nature of injustice and the desirability of punishment as the cure. This proposal is not unprecedented, for it has been anticipated by the Stranger’s earlier claim that crime is rooted in the weakness of human nature. By this proposal the Stranger makes it clearer than he had before that his understanding of crime rules out retribution as a proper response to crime.

² Cf. Strauss, *The Argument and the Action of Plato’s ‘Laws,’* 128-29.

Kleinias' reaction to this proposal, however, still leaves it uncertain whether or not he has noticed its implication for retributive punishment. Kleinias expresses his agreement with the scolding slave doctor conjured up by the Stranger (857e). Apparently, Kleinias does not see any need for a medical patient to learn about the nature of his pathology and the effects of the cure if his real interest is only in becoming healthy, and if a doctor can produce health just as easily with the patient's full knowledge of the nature of the treatment as without it. For Kleinias, the same appears to apply to punishment. If the purpose of legislation is to make men obey the laws, and if this can be accomplished just as easily by threatening them with punishment as by educating them (if not more easily), then the Stranger's proposed educational approach to crime would be superfluous at best. If, contrary to the Stranger's assertion, the free patient has no need to learn about the nature of bodies to be healthy, then by analogy the criminal would require no special knowledge about the nature of the soul in order to become just again, according to the view Kleinias seems to have adopted. Although this is an important insight into Kleinias' view of the function of punishment, we still do not have adequate information to determine what exactly he thinks of retribution, since it is still possible that he was reluctant to accept the Stranger's educational model of punishment not only because he thinks it would be superfluous, but also because he thinks it would be too lenient on criminals, and thus less just. The Stranger's depiction of the slave doctor's reaction to the free doctor as indignant (*Ō mōre!*) may be intended to remind us of the second alternative.

The Stranger next suggests that, contrary to Kleinias' assumption, legislation need not interfere with education, and that somehow punishment and education may be

combined (857e). Perhaps the Stranger is thinking of introducing preludes that would expand on the arguments of the earlier public penal prelude in order to show more clearly how injustice in the soul harms the unjust man, and how punishment would remedy his condition. The Stranger now observes to his companions that they are in a fortunate situation since they are not under any necessity to legislate immediately, but by becoming engaged in a leisured inquiry they might “try to discern in what way the best and also the most necessary would each come into being.” “Indeed,” he continues, “it is likely that we can now, if we wish, inquire into what is best, or, if we wish, into what is most necessary, as regards the laws—so let’s choose which it seems we’ll do” (857e-858a). The Stranger thus assures Kleinias that combining education and punishment would be a finer thing than merely relying on the latter, and holds out the promise that he can show Kleinias a way of combining the two. We note that the Stranger thereby concedes that the arguments of the penal prelude in Book V were insufficient to educate citizens about the nature of injustice in the soul and about the value of punishment as a cure. Being a legislator, Kleinias’ love of glory is sensitive to the praise or blame his laws receive from posterity. Thus when the Stranger offers him the choice between laying down laws that are merely “necessary” and those which are “best,” Kleinias enthusiastically chooses the latter (858a-c). In this way, by appealing to Kleinias’ love of glory, the Stranger succeeds in inviting him to participate in “a precise inquiry” into the nature of “all injustices” (859b-c).

The Disharmony of the Legislators

Having won Kleinias' confidence, the Stranger recommends that they proceed to examine the opinions of the famous Greek legislators. The Stranger argues that those who wish to be happy would do well to study any of the writings that may have been left behind by the great legislators—especially Lycurgus and Solon (858e)—about the noble, the good, and the just things. But examining the opinions of the celebrated Greek legislators is not quite the same thing as examining what is best concerning legislation, which is what the Stranger had actually proposed he and Kleinias should do. The Stranger appears to assume that the famous legislators had left behind them written preludes explaining the wisdom of their laws, a possibility he ruled out moments earlier when he asserted that human beings living under *all actual legislation* live as slaves being doctored by slaves (cf. again 857c).

We are further surprised to find that instead of investigating what is best concerning legislation, which is what the Stranger initially proposed, and instead of inquiring about the noble and the just things from the writings of the famous Greek legislators, which he suggested in his revised proposal, the Stranger actually proceeds to examine how these things stand in the opinion of the interlocutors themselves, that is, “to what extent we now agree, and to what extent we disagree with ourselves” regarding “the noble things and all the just things” (859c-d). Furthermore, since Megillus never speaks in the ensuing exchange, and since Kleinias does all the answering while the Stranger asks questions, it is really only Kleinias' opinions about the noble, the good, and the just that are subjected to examination. It is not clear why the Stranger proceeds in this

puzzling way, particularly, why he wishes to focus on what is sound and what is unsound in Kleinias' opinions about the noble things and the just things. Perhaps the Stranger means to imply that Kleinias' opinions about these matters will be representative of the opinions of the great legislators because he thinks that their opinions about these matters are essentially the same. Since the great legislators had not in fact left any prelates to explain the wisdom of their laws, the best access we have to their opinions is through Kleinias.

Even if this interpretation of the Stranger's puzzling way of proceeding here is correct, it leaves us to wonder about its relevance to his purpose of combining legislation with education. The Stranger had stated unambiguously that he thinks nothing that pertains to laying down laws had been correctly worked out in any way, which would mean that the most that could be hoped for from a critical examination of the opinions of past or present legislators is greater awareness of their errors, whatever those errors might be. But how would this information help men suffering from injustice in their souls better understand the cause or nature of their suffering, and how would it help them accept punishment as the necessary remedy? Does the Stranger believe that knowledge of the nature of injustice in the soul can somehow be gleaned from knowledge of the errors of the legislators about the noble, the good, and the just things? Could knowledge of these errors somehow be the truest prelude to the penal laws?

The Stranger begins to lay bare Kleinias' deepest beliefs regarding punishment by asking whether "our" opinions about the just and the noble are consistent. He elicits Kleinias' assent to the following:

With respect to justice as a whole and just human beings, deeds, and actions, we all somehow agree that all these are noble; so if someone would maintain that just human beings, even if they happen to be ugly in their bodies [*aischroi ta somata*], were nevertheless entirely beautiful people in respect to their very just disposition considered by itself [*kat' auto ge to dikaiotaton ethos tautei pankalous einai*], in almost no case would the one who speaks thus seem to speak off key. (859d-e)

The Stranger proceeds to argue, again with Kleinias' assent, that if all things are noble which partake of justice, then this would include the things we undergo or suffer as much as the things we do. But if this is the case, the Stranger goes on, then they have contradicted the penal laws they had established earlier, for “presumably we established that the temple robber should die—justly—and the same for the enemy of well-made laws . . . and that while they were the most just of all sufferings, they were also the most shameful [*aischista*]” (860b). In this way, the just things and the noble things appear “to us” at one time to be all the same, and at another time opposed (*ibid.*). The Stranger adds that this inconsistency or confusion exists not only among “us” but also among “the many,” who, in view of these things, “proclaim without consonance that the noble things and the just things are separate” (860c). On the question of the identity of the just things and the noble things, the opinion of the legislators and the opinion of the many are essentially the same.

Kleinias thinks he is guilty of the contradiction of which he is accused by the Stranger (860c). But was this contradiction in fact necessary?³ In the first place, notwithstanding what was said in the public penal prelude, is it really necessary to define justice, and thus also just punishment, as something that is entirely noble, or isn't it the case that some things are just without also being noble? For example, the law commands some things because they are noble and choiceworthy for their own sake (such as education), but it also commands other things merely because they are necessary (such as sewage disposal). In this way, while all of the things commanded by the law would be just, not all would be noble. Could we not then regard punishment as one of the things that are just but not noble? Alternatively, the Stranger had just observed that we commonly distinguish between the beauty of the soul and that of the body. Thus, even if punishment of criminals damages their bodies and makes them physically ugly as a result of their suffering, it might still be argued that such suffering affects their souls in the opposite way: by undergoing just punishment their souls become better (as has been supposed throughout the discussion) and therefore more beautiful or noble in this more important respect. The Stranger virtually hands Kleinias the means by which to escape his self-contradiction, leaving the reader wondering why Kleinias does not avail himself of it. At the very least one might argue, as the Stranger had earlier, that if the unjust man's character becomes so incurably bad that death would be the lesser evil for him, then executing him would be a noble thing both for himself and for the city.

³ Cf. Leo Strauss, *Plato's 'Laws,'* 130.

The fact that Kleinias does not take refuge in the second alternative (arguing that punishment is nobly suffered by the unjust man because he is improved) is highly curious, and prompts us to wonder about its cause. Could it be that Kleinias is reluctant to agree that suffering just punishment is noble because he *wants* punishment to be harmful to the criminal in every way, and in no way beneficial? It seems that the Stranger has finally provoked Kleinias into revealing his sincere belief that retributive punishment, which harms rather than benefits criminals, is just.⁴ Whatever else it meant, Kleinias' objection to the Stranger's original law regarding the punishment of theft now seems to have been closely connected with his belief in the justice of retribution.

What, then, can we say about the other horn of Kleinias' self-contradiction: why does he not attempt to escape his inconsistency by using the other alternative I mentioned, namely, to argue that some instances of justice are necessary but not noble and that suffering just punishment is one of these? I believe the reason for this can be discerned from Kleinias' earlier enthusiasm for the Stranger's proposal to make their laws look to what is best rather than to what is merely necessary. As a politically ambitious and glory-seeking man, Kleinias desires to be praised as a great benefactor of human beings. The Stranger had convinced Kleinias that in order to earn such praise his legislation would have to look beyond addressing the problems of crime, faction, and foreign aggression to cultivating virtue as a whole, which, as the Stranger had hinted, could only be

⁴ This may explain the reason for the Stranger's implausible argument that actions are always suffered in the same manner as they are performed. It now appears that this was a device to allow the Stranger to draw the conclusion that suffering punishment ought to be noble, thus putting Kleinias in a position where he would have the opportunity to deny this conclusion, thus revealing something about himself. On these grounds I disagree with Saunders' interpretation of this passage as merely a preparatory "lesson in predication." See Trevor Saunders, "The Socratic Paradoxes in Plato's *Laws*: A Commentary on 859c-864b," 422.

accomplished by providing for a rational education as to the goodness of justice for the just man's soul. Since, as the Stranger had claimed, this had never been achieved by any previous legislator, not even by the famous Lycurgus and Solon, Kleinias realizes that he has an opportunity to be honored as the greatest Greek legislator ever.

It has now become evident that Kleinias is not, and has never been, simply motivated by his political ambition alone, for we now know that this motivation exists alongside, and in tension with, a concern with retribution of which he is hardly aware. Even when the Stranger had explicitly brought to Kleinias' attention the contradiction implicit in his view of justice—namely, that justice appears to him at times to be noble (i.e., beneficial) and at other times not noble (i.e., harmful)—he remained unwilling to abandon his conviction that just punishment must harm rather than benefit the criminal. If Kleinias had been aware of his genuine commitment to retributive justice, then he surely would have objected long ago to the Stranger's definition of crime as analogous to disease, and to his definition of just punishment as the cure. Kleinias had not objected earlier because he had not noticed the full implications of the Stranger's position, which was due at least in part to the Stranger's equivocal use of the word *timōria*.

If we reflect further about Kleinias' opinions regarding punishment, we can see that the confusion the Stranger has uncovered has further and even more serious personal consequences for Kleinias. The Stranger's analogy between crime and disease, and between legislation and medicine, implied that as the aim of medicine is to make men healthier, the aim of legislation is to make them more just. If medicine should cure diseases of the body by means of beneficial treatments, then legislation should cure

injustice in the soul by way of beneficial punishments, making the soul healthier by making it more just. Kleinias insisted that undergoing punishment must be harmful rather than beneficial to the criminal, which contradicted the analogy between punishment and cure. At the same time, while Kleinias had denied the *beneficial* effect of punishment, he never repudiates his view that punishment makes men more obedient to the laws, and thus *more just*. But if making the criminal more just through punishment ultimately harms him, as Kleinias insists by denying that suffering just punishment is noble, then how can the virtue of justice in general still be thought to be the good condition of the soul? To repeat, Kleinias had initially agreed with the Stranger that regardless of a man's physical condition he is entirely noble to the extent that his soul is just, since justice is the health of the soul. By later denying that suffering just punishment (and thus becoming more just) is noble, Kleinias implicitly denied that justice is the health of the soul. And yet, Kleinias does not unequivocally deny that justice is the good condition of the soul either, since he is equally committed to both sides of his self-contradiction. Rather, he seems to believe both of these things to be true at the same time—that justice is the good condition of the soul and that it is not. It goes without saying that both of these things cannot be true, and so we are forced to conclude that Kleinias wavers or has doubts about the goodness of justice.⁵

⁵ One potential difficulty with my interpretation of 860b is that while I draw from it the conclusion that Kleinias is confused or doubtful about the *goodness* of justice, the contradiction pointed out by the Stranger is explicitly said to be only about Kleinias' view of the *nobility* of justice. The reason why I think that Kleinias' view about the goodness of justice is implicated as well is that, as an intensely political man, Kleinias identifies the noble with the highest good. The noble is the highest good from his point of view because it is not only beneficial to oneself (i.e., honorable) but also beneficial to others (cf. again 728b-c). Consequently, I believe that the Stranger chooses to address Kleinias' confusion about the goodness of justice in terms of the noble because the noble is that which Kleinias values most highly.

Kleinias' doubt about the goodness of justice, as revealed through his opinions about punishment, is one instance of a psychic phenomenon that plays a very important role in Plato's dialogues. Plato's Socrates was unmatched in his ability to win arguments on almost any topic, and against any sort of opponent, not only because of his manifestly powerful intellect and mastery of rhetoric but also, and more importantly, because of his greater awareness of his opponents' ignorance about the "noble and good" things (cf. *Apology* 21b-22e). Through his many conversations with men in all walks of life, Socrates had learned that the greatest obstacle preventing most people from gaining a greater awareness of their ignorance about the most important things is not their indifference about these things, as it might initially appear. Rather, it is precisely the power that these things hold over men's hearts that prevents them from subjecting their conflicted opinions about these things to thorough examination for fear that such an examination might lead to the disappointment of their hopes. Thus most men, according to Socrates' view, remain deeply conflicted throughout their lives about justice, the noble, and the good, while only a few, like Socrates himself, are able to acknowledge to themselves the seriousness of their condition, even if only for a few brief moments in their lives (cf. *Alcibiades I* 116e-118b).⁶ This view is echoed by the Stranger in the passage we are presently considering when he attributes the confusion about the identity of the just and the noble things not only to themselves, as legislators, but to "the many" as well—that is, to all political men.

⁶ See also Christopher Bruell, *On the Socratic Education*, 97-102.

I would like to consider a few possible implications of this deep-seated doubt about the goodness of justice as it applies to Kleinias. Kleinias had agreed that justice is the good condition of the soul and that injustice is the greatest evil that can afflict it, but he had initially objected to the Stranger's proposal that they should include preludes for their penal laws following the model of the free doctor who philosophizes with his patients. He had objected to this because he believes rational persuasion to be unnecessary or superfluous for making men just. Evidently, Kleinias believes that what makes men just is the cultivation of the habit of obedience in combination with the threat of punishment. Perhaps he tends to think that the virtue of justice itself is nothing more than the habit of lawfulness. But would men not be at least somewhat more inclined to be just by the rational knowledge of the goodness of justice than they would by habituation and fear of punishment alone? Why, then, was Kleinias so initially uninterested in educating criminals philosophically, so much so that the Stranger could only procure his assent to this approach by appealing to his love of glory?

Part of the answer begins to emerge when we consider this question in light of what we have just discovered—that is, in light of Kleinias' wavering regarding the goodness of justice. When the virtue of justice is understood as a habit of lawfulness, as Kleinias seems to understand it, it is difficult, if not impossible, to think of it as good in any sense other than as producing lawful actions, which are beneficial because they contribute to the general safety and good order. As such, the virtue of justice becomes difficult, if not impossible, to defend as intrinsically good (good for the soul), rather than as merely good for its secondary consequences, such as safety and a good reputation.

From this point of view, the Stranger's proposal to educate criminals philosophically as to the goodness of justice for their souls would seem ill-advised.

What is more, Kleinias does not show interest even in trying to encourage law-abidingness by teaching that justice is good because of its consequences. Not unlike Hobbes, Kleinias fears that the truth about human relations may be that, on every level, there exists a war of all against all, a condition in which true goodness consists in advancing one's own interest without any consideration for the interest of others (cf. 626b-d). On this view, just as it would be in accordance with nature for cities to dominate other cities rather than to treat them justly, it would be equally in accordance with nature for individuals to exploit each other by force and fraud. Justice, then, would be at best a conditional (and thus fragile) pact, or social contract, among individuals to protect each other against common enemies. At worst, justice would be a kind of fraud that some men practice on others in order to make them more willing to serve their interests while neglecting their own. Kleinias dearly hopes that this is not the case (cf. 627b-c), and yet he does not think he can prove to himself, or to others, that it is not. In short, Kleinias had been initially reluctant to combine his legislation with philosophical preludes on the subject of punishment, according to the Stranger's proposal, because, unlike the optimistic Hobbes, Kleinias did not believe that human beings can be persuaded through rational argument that justice is *always* advantageous for them individually, whether as an intrinsic good or as an instrumental one. Kleinias fears that a philosophic education about justice would simply not work, or worse, that it might have the contrary effect.

As I argued earlier, the Stranger's odd procedure in this section of the dialogue seemed to suggest that Kleinias' opinions regarding the noble, the good, and the just things would be representative of the opinions of the famous Greek legislators—and perhaps of all political men. I also suggested that the Stranger intended for the examination of the legislators' errors about these matters to serve, somehow, as the truest prelude to the penal laws. I believe we can now summarize how that examination was meant to function in this role. The goodness of justice had been taken for granted throughout the dialogue and, moreover, it was more or less assumed that evidence for this could be provided to the whole city by the legislators through civic education, and specifically through educational preludes. It has now become clear that according to the legislators themselves (insofar as Kleinias represents their views) no such evidence is readily available. At the beginning of the discussion in Book IX, the Stranger had made the questionable claim that all crime is due to the weakness of human nature, and specifically to its inborn tendency to resist civic education. We now see that the problem lies not in human nature but in civic education itself—i.e., in civic education as such. By implying that the examination of the opinions of the legislators, which revealed their doubts about the goodness of justice, would itself serve as the prelude to the penal laws, the Stranger may have been suggesting that it is the truest explanation for why penal laws are necessary. Penal laws would not be necessary if the fundamental cause of crime could be removed, but the fundamental cause of crime cannot be removed by means of legislation, since legislation is not sufficiently able to persuade citizens that injustice really is the greatest evil that can afflict the soul, but only partially by means of

admonitions and threats. As we saw, at 859d, and later at 860c, the Stranger attributes the confusion about justice not only to the legislators, the virtuous few, but also to “the many”—that is, to all political men. Thus the Stranger was right when, at the beginning of the discussion in Book IX, he said that he had a valid excuse for laying down penal laws even though this was in a sense shameful (i.e., necessary but not noble), except that he had then misrepresented what that excuse really was. The city’s assumption that the intrinsic goodness of justice is either self-evident or that it is teachable through civic education is now revealed to be only a hope. The fundamental cause of crime is, therefore, not human nature’s resistance to education, but the city’s inability to provide an adequate education with respect to justice as a result of its own profound doubt regarding its goodness for the individual.

In arguing that legislation is not capable of removing the fundamental cause of crime, I do not mean to imply that this cause could not be removed in other ways. Could the cause of crime in the soul be removed through philosophy or through prayer? In the *Laws*, Plato does not have the Stranger attempt to purge the cause of criminality from the souls of his interlocutors by means of philosophy for the obvious reason that they are too set in their ways and too unfamiliar with philosophy to make the attempt. It goes without saying that to purge the cause of criminality by means of philosophy from the whole city would be impossible. And while the Stranger does not seriously attempt to purge the cause of crime through prayer (but see 854b-c), he does attempt to provide an alternative justification or prelude for the penal laws by way of a theological discussion in Book X that claims to prove the existence of just and avenging gods, and which constitutes one of

the peaks of the *Laws* as a whole. Indeed, in the context of that theological discussion, Kleinias—though not the Stranger himself—calls that discussion “the noblest and best prelude” on behalf of all the laws (887b-c).

In connection with this alternative, theological approach to the question of punishment, I would like to suggest that the Stranger’s provocation of Kleinias’ genuine concern for retributive justice also provokes and brings to life his latent hopes for, and embattled belief in, divine providence (cf. 624a3-4).⁷ The public penal prelude had encouraged citizens to be spirited in their outlook on crime, but the connection between spiritedness and piety was left unexamined. While this connection is not examined in Book IX either, the lengthy treatment of the sacrilegious crime of temple robbery prepares Kleinias to react to the Stranger’s legislation for the punishment of theft by raising the issue of retributive justice. We can infer from these observations that Plato wished to suggest that the question of retributive punishment is intimately bound up with the question of divine providence, and that it cannot be properly understood without the latter.⁸ It is in this part of the dialogue, when Kleinias’ belief in the justice of retribution is directly challenged (as we shall see shortly), that the dialogue begins its ascent to the question of the existence of the gods and of their justice, culminating in the Stranger’s conversation with the impious youths in Book X.

⁷ Cf. Strauss, *Plato’s ‘Laws,’* 3.

⁸ For a discussion of the connection between punishment and piety in the *Laws*, see Thomas Pangle’s “The Political Psychology of Religion in Plato’s *Laws*.”

Voluntary and Involuntary Injustice

Kleinias is not very happy about having had his confusion about justice exposed, but has nothing to say in defense of his self-contradictory position, and so the Stranger gracefully comes to his aid. The Stranger now proposes that they look together to see how they, as distinguished from the many, have a harmonious view of the just and the noble, despite their apparent confusion (860c).

When Kleinias agrees to let the Stranger help him out of his bind, the latter goes on to state a view which he claims to have expressed earlier, “that the bad are all bad involuntarily in every respect” (860d). He explains:

The unjust man is presumably bad, but the bad man is involuntarily so. Now it never makes sense that the voluntary is done involuntarily. Hence the man who does injustice appears involuntarily unjust to the one who sets down injustice as something involuntary. This is what I must agree to now. For I agree that everyone does injustice involuntarily. And if someone, out of love of victory or love of honor, asserts that the unjust are indeed involuntarily so, but that many voluntarily do injustice, my argument, at any rate, remains the former and not the latter. (860d-e)

This view had already been implicit in the Stranger’s account of injustice as a disease in the soul. No one would wish to become sick, and so if injustice is a disease of the soul then anyone who becomes unjust does so involuntarily. It will be recalled that the Stranger had in fact already stated this very thesis in the public penal prelude in Book V,

although there he had limited its scope to curable injustices only (cf. 731c). The incurable criminals, those who are thoroughly evil, were assumed to be unjust voluntarily, since it was then said that they are worthy of anger (cf. 731d). As we observed in our earlier discussion of that passage, the prelude gave no basis for its claim that incurable injustices are voluntary. If disease is not something we ever desire, why would it make a difference if it is curable or not? Perhaps what the public prelude meant was that in cases of incurable injustice the will or desire itself is corrupted, so that what the incurably unjust man takes pleasure in, and what he really desires, *is* this disease of the soul.⁹ In this way, incurable injustice may be thought to be voluntary. What, then, has prompted the Stranger to change his mind now and say that *all* injustice is involuntary?

The Stranger gives no arguments to substantiate his radical thesis, but says that he must maintain it because he believes it to be true, and because it would not be pious or according to custom for him to lie (861d). Does the intervening discussion of the opinions of the legislators, which we have been following closely, somehow provide the arguments to substantiate the Stranger's radical claim? The most important thing we have learned from the Stranger's examination of Kleinias' opinions about punishment is that the fundamental cause of crime is the doubt about the goodness of justice which defines the political outlook. Someone in doubt about the goodness of justice vacillates, thinking at one time that justice is better than injustice and at other times that injustice is better (cf. *Lesser Hippias* 372d-e). But isn't doubting something the same as being in ignorance about it, and don't we believe that those who act in ignorance act involuntarily, if to make

⁹ Cf. Aristotle, *Nicomachean Ethics* 1113a16-b2.

a free choice is to make it with full knowledge of what one is doing? Contrary to the message of the public penal prelude, no one who acts unjustly desires injustice wholeheartedly and unequivocally—no one is “thoroughly evil” if all are in doubt about the value of justice. Even those who profess to be most contemptuous of justice become angry, thus revealing a deeper concern for justice (cf. *Republic* 336b; *Gorgias* 484a-b, 485b-c). In this new light, no one who does injustice can be said to do it voluntarily. The Stranger’s view that the fundamental cause of injustice is ignorance about the goodness of justice destroys the distinction between the curably and the incurably unjust men. The man who is “cured” of his injustice by admonition, habituation, or threats is not cured of his ignorance, but is persuaded only for a time to refrain from injustice (cf. 854b-d). In light of these far-reaching implications of the examination of Kleinias’ opinions, the Stranger can now say that all injustice is involuntary, since all injustice is done in ignorance.

The political consequences of the Stranger’s view of injustice, now made perfectly explicit, can no longer be ignored: all punishment that is not educational would not be appropriate because any punishment that does not aim to cure the ignorance which is at the root of crime provides no hope of real improvement, whereas justice has been assumed to be beneficial, and because involuntary injustice cannot be said to deserve retribution. Non-educational punishment may still be justified insofar as it deters the individual from committing further injustice, thereby harming himself and his city, and inasmuch as civic education is poorly equipped to correct the ignorance at the root of crime (which, as I have argued above, is the Stranger’s truest justification of his penal

code). Such punishment would, however, be dictated by necessity, and thus could not be noble or just in the highest sense. But more importantly, and to repeat, the Stranger's radical thesis completely rules out retributive punishment, and this realization has now dawned on Kleinias for the first time. The Stranger now explicitly puts the question to himself and his two interlocutors how his new understanding of injustice would affect their legislation for the city. Would they have to distinguish the involuntary injustices from the voluntary, establishing greater penalties for the voluntary crimes and lesser for the involuntary, or equal penalties for all on the grounds that there are no voluntary injustices at all (860e-861a)?

At this stage in the conversation, Kleinias finds himself in a highly frustrating position. He began by being dissatisfied with the Stranger's overly simple approach to punishing theft and demanded that the penal laws of the city reflect the qualitative moral differences among crimes. The Stranger responded to this objection by comparing all actual legislation—including the oldest and most revered—to the despotic commands of slave-doctors to their fellow slaves. Then, just as the Stranger seemed to hold out the promise of transforming the laws of Kleinias' new city into the best and most praiseworthy laws, thereby winning immortal glory for Kleinias as the father of those laws, the Stranger instead went on to mount an even more devastating critique of actual legislation, and of Kleinias himself, by bringing into question the soundness of the conventional understanding of justice. Finally, in lieu of providing the promised, graceful escape from Kleinias' embarrassment, the Stranger further attacked the conventional

view of punishment, accepted by Kleinias himself, by undercutting one of its basic assumptions: the distinction between voluntary and involuntary injustice.

In passing, the Stranger hints at a way in which Kleinias could still escape his present difficulty and maintain the conventional distinction between voluntary and involuntary injustices. The Stranger does this by raising the possibility that the distinction between voluntary and involuntary injustice may have been “spoken by a god, giving no argument as to why it has spoken correctly,” but then quickly rejects that possibility without supporting arguments and without giving Kleinias the opportunity to respond (861b-c). In this way, the Stranger points to the serious alternative of justifying the conventional view of punishment on the basis of divine revelation, an alternative which he considers at length in Book X. The Stranger avoids expressly engaging in such a discussion at this stage, however, and instead goes on to elaborate a public doctrine of criminality that dimly reflects the philosophical insight which he has caused to emerge, while superficially accommodating the conventional view of punishment.

The Stranger's Twofold Conception of Voluntary Action

Kleinias' state of helpless consternation leaves him with no choice but to allow the Stranger to follow his present course in whatever manner he wishes so that they might bring their theoretical discussion of crime to a conclusion and continue with the legislation. The Stranger now proposes that they make a distinction among crimes on some other basis than the difference between the voluntary and the involuntary. In this way they would resolve their difficulty and make the practice of punishment intelligible

enough so that “everyone may follow” whenever a judicial penalty is imposed (861c-d). He signals thereby that what comes after will be more like an edifying public prelude than what has preceded. It also bears keeping in mind that this new turn in the conversation is part of the Stranger’s attempt to show how the legislators (he and Kleinias), as distinguished from the many, are harmonious in their views about justice even though the self-contradictions revealed in the cross-examination of Kleinias have been left unresolved. What follows is a politically salutary account of the psychological causes of crime which attempts to accommodate the ordinary public-spirited outlook on crime while at the same time gently moving that outlook as much as possible in the direction of the Stranger’s enlightened understanding.

Speaking on behalf of the legislators, the Stranger now says that some harms (*blabai*) committed by citizens against one another in their associations have the character of the voluntary (*hekousion*), while others have the character of the involuntary (*akousion*). But harm is not as such an injustice, whether it is great or small, nor is benefit justice. Rather, “when a benefit comes to pass that is not correct, the one responsible for the benefit is committing an injustice” (862a-b). Generally speaking, “if someone gives something to somebody, or, on the contrary takes something away, such a thing shouldn’t thus be called simply just or unjust, but what the lawgiver should observe is whether someone employs a just disposition and character [*ean ethei kai dikaioi tropoi chromenotis*] in doing some benefit or injury to somebody” (862b). The lawgiver should concern himself with both injustice and injury, but he should keep the two strictly separate in his legislation. He should do what he can through his laws to redress the injury, and then

once compensation has been made he should strive to create “friendship in place of discord” among the citizens (862b-c). Compensation for any damage done by crime has nothing to do with the injustice of the crime, and does not take the motivations of crime into account; the sole concern of compensation or restoration is with the good.

What has been damaged must be restored, but the unjust character responsible for the crime must be “cured.” By curing, the Stranger appears to mean not only punishment but something much broader: in curing injustice the legislator should “teach and compel” the unjust man “either never again to dare voluntarily to do such a thing or to do it very much less” (862c-d). To accomplish this, the legislator should employ deeds or words, pleasures or pains, honors or dishonors, and even fines or gifts. By these means he should try to “bring about hatred of injustice and desire, or lack of hatred, for the nature of the just” (862d-e). As for the offender whom the lawgiver perceives to be incurable, he should institute a judicial penalty (*dikēn*) of death, because he knows, “presumably,” that it is worse for him to go on living. Such capital punishment benefits the offender himself by imposing the lesser evil of death, and at the same time confers a “double benefit” by setting an example for would-be criminals and by emptying the city of bad men (862e).

The passage I have summarized in the two previous paragraphs is somewhat perplexing, and its interpretation is in dispute among commentators. The source of the perplexity is the Stranger’s ambiguous or equivocal usage of the term “voluntary.” In the first part of the passage, the Stranger maintains that when we benefit or injure others we do so voluntarily. By this he appears to mean that we try to benefit ourselves and our friends and to bring harm on our enemies. In this way, the Stranger seems to link

voluntary action with the pursuit of one's own good, in accordance with his stated thesis that no one ever chooses what is bad for oneself voluntarily. But on the other hand, the Stranger also says in this passage that the legislator should attempt to teach the unjust man never again to do injustice "voluntarily." Initially, this makes little sense because justice and injustice were defined by the Stranger as dispositions of character considered separately from the consequences of our actions. We are compelled to ponder the possibility that the Stranger is using the term voluntary in different senses when he is referring to actions done for the sake of some good or advantage on the one hand, and when he is referring to just and unjust actions on the other. Commentators explain the Stranger's twofold view of the voluntary as involving a tacit distinction between voluntary action in a merely "legal sense"—that is, as freedom from external compulsion to act in accordance with the law or against it—and voluntary action in the full or true sense, grounded in a comprehensive understanding of the ultimate objects or ends of human action.¹⁰

What, then, is the import of the "legal sense" of voluntary action? According to this sense, our choices are to be judged in reference to our characters—i.e., whether they are just or unjust—wholly independently of considerations of any benefit or harm resulting from those choices to particular individuals or to the city. If the legal definition were followed to its logical conclusion it would mean that acting justly is neutral towards the good—it is neither good nor bad necessarily. In this way, the "legal sense" of voluntary action may be said to divorce justice from the good, contrary to the Stranger's

¹⁰ See Mackenzie, *Plato on Punishment*, 201-202; Saunders, "The Socratic Paradoxes in Plato's *Laws*," 422-424; Strauss, *Plato's 'Laws'*, 132.

thesis which had implied that we can only choose that which we know to be good for us.¹¹ Neither of the Stranger's companions protests his departure from his earlier thesis, perhaps because they are not as invested as the Stranger appears to be in reconciling the penal code with the implications of that thesis. But perhaps neither protests for another reason as well: while the legal definition of voluntary action is incompatible with the Stranger's thesis, it is not an entirely foreign concept to Kleinias (or to ourselves). It is commonly said that one has a duty to be just whether or not it is in one's own interests, that is, regardless of the rewards or sacrifices involved in being just. This is illustrated by Kant's dictum that justice is prior to the good. The "legal sense" of voluntary action finds purchase with Kleinias because of his agreement with conventional opinion.

Yet the Stranger's foregoing discussion has prepared his more attentive audience (the highest magistrates among them) to see the deeper significance of the legal definition of voluntary action. We can now see that the legal definition echoes what the Stranger has revealed to be the city's doubting hopefulness regarding the goodness of justice. Political men waver regarding the relative value of justice and injustice, but they sincerely wish to affirm the choiceworthiness of justice in spite of their wavering. The civic-minded belief that one ought to do one's duty for its own sake, and not for the sake of a reward, expresses this affirmation of the choiceworthiness of justice perfectly. But it does more than just express this civic attitude; the legal definition of voluntary action also appears to give it authoritative support and justification. According to the legal definition, one does not need to take pains to know whether justice is good for the just man or

¹¹ Cf. Strauss, *Plato's 'Laws,'* 131-33.

whether injustice is in fact better for him, for justice is choiceworthy for its own sake regardless of whether or not it is good. But this justification is only apparent, because, as we have seen, the outlook that defines political life wishes for justice to be good and choiceworthy *because of its goodness*, which is very different from saying that justice is *prior* or indifferent to the good. It will be recalled that Kleinias refused to retreat from his view of justice as entirely beneficial in every way. The legal definition of voluntary action therefore misrepresents the political outlook even as it gives it authoritative sanction.

The Stranger's account juxtaposes the legal definition of voluntary action, and even allows it to be confused with, what I have called the true definition of voluntary action, which identifies voluntary action with the pursuit of the good. The reason why I have called it the true definition is because it is the one accepted by the Stranger (cf. again 860d), and because our discussion of the Stranger's examination of the opinions of the legislators has given us reason to believe that the laws themselves aim at the good, as does the virtue of justice (cf. *Minos* 314c-d; *Republic* 331c-d). The Stranger is compelled to present the true conception of voluntary action in tandem with the legal definition not just because he knows the former to be true, but also because it is required by the principle established in the public prelude of Book V, that justice is beneficial and not harmful. The Stranger regards both conceptions of voluntary action as politically necessary. In accordance with the true conception, citizens will believe that by being just they pursue their own and the city's good. On the other hand, since doing one's legally prescribed duty occasionally requires restricting the pursuit of one's private interests, and

sometimes even the sacrifice of some private good, the legal conception of voluntary action allows the city to hold individuals responsible for obeying the law. Although these two definitions of voluntary action remain theoretically incompatible, in practice they combine to form a workable doctrine of legal obligation.¹²

I do not wish to leave the impression that the Stranger's twofold doctrine of voluntary action merely repeats the conventional view of legal obligation, crime, and punishment which he in fact regards as confused. While respecting the essential limitations of the political outlook, the Stranger attempts to curb the city's tendency towards retributivism by infusing his discussion with the language of medicine, which he hopes will have a moderating and humanizing influence on the public administration of punishment by making the city associate injustice with disease (862c8-9). In addition, the Stranger tries to moderate the city's tendency to sacrifice the individual to its own interests by maintaining that justice must be noble in every respect, that is, beneficial to the city and to the individual at the same time (862d10, e4). In this way, the Stranger hopes to temper somewhat the city's preoccupation with its own security and to encourage it to balance the need to deter crime with its care for the well-being of its individual members. But unlike modern political philosophers, the Stranger does not attempt a wholesale overhaul of the city's outlook on crime, since he understands that outlook too be determined by the horizons of political life and required by the basic needs of a functioning society.

¹² I believe that this helps explain the failure of Kant's attempt to make them compatible through the conceit of the *summum bonum*.

It is true that as certain forms of political society—such as our own liberal democracies—relax the need for a spirited and honor-loving citizenry, the humane view of crime is able to gain greater popular influence, and this is something for which liberal societies ought to be praised. At the same time, I would suggest that we should not let the differences between the city of the *Laws* and our own liberal democracies obscure the merits of the Stranger’s teaching about what is universal to political society as such. It is arguable that even in liberal republics, where reasonableness is encouraged, where spirited indignation is regarded with suspicion, and where the fruits of science and philosophy are widely disseminated, spiritedness will continue to exist as an essential ingredient of political life, as such. If Plato is right, then perhaps even a liberal society must make room for the retributive outlook, despite all of its tensions.

The Popular Psychology of Crime and the Completion of the Penal Legislation

The theoretical section of Book IX is concluded with the Stranger elaborating a popular psychology of crime (or error: *hamartema*). Instead of the most fundamental cause of crime, to which he had quietly called attention earlier, the Stranger now describes three different causes of crime in the soul that are recognized by popular opinion (863b: “you say and hear from one another”). The first two of these causes are spiritedness and pleasure, either of which can urge a man to act contrary to his choice or intention (*boulēsis*). Pleasure can have this effect on one’s decisions “through persuasion and forceful trickery,” while spiritedness does so by inciting one to be “quarrelsome,” “combative,” and “uncalculating” (863b; cf. 863e). Yet it is also commonly believed that

we are ultimately in control of our passions since we can be either “weaker” or “stronger” than them (863d). The third cause of crimes is ignorance, which is further subdivided into two forms: the first version is responsible for light and forgivable errors, while the second version occurs when someone not only lacks understanding but also believes that he is wise. Penal laws established for crimes caused by the simple form of ignorance will be gentle and forgiving, but those established for crimes stemming from presumptuous ignorance will not. This third cause, too, directs man to act against his intention, but unlike the other two causes, it is not commonly believed that anyone can be “stronger” or “weaker” than his ignorance (863d). It seems to be implied that whereas we are capable of overcoming our passions through reason, discipline, or force of character, we are helpless when it comes to ignorance.

By distinguishing between ignorance and the passions of desire and spiritedness as causes of crime, this criminological scheme assumes that wrongs committed out of desire or spiritedness may be committed with knowledge of what one is doing, and thus voluntarily, if one was able to control one’s passions but did not. This entails that crimes committed for the sake of pleasure or out of spiritedness can be worthy of blame and possibly even of retribution. In this way, the Stranger appears to keep the promise he had made to Kleinias earlier that he would show how the conventional opinions of the legislators are harmonious after all. He does so now by describing how some injustices can be voluntary and others involuntary, thus preserving the important conventional legal distinction between voluntary and involuntary injustice. He is not deterred by the fact that

in doing so he contradicts what he himself believes to be true: that all injustice is involuntary.

Since crimes due to ignorance accompanied by the opinion of one's wisdom are said to be unforgivable, the Stranger seems to be implying that we can be held responsible for our ignorance as well, at least in this limited category of cases. In this, too, the Stranger seems to be at odds with what he had indicated at earlier stages in the conversation. The Stranger now says that he can define what the just and unjust are "without complication." The tyranny in the soul of the passions, whether it does some injury or not, is injustice, while the rule of "opinion about what is best, however a city or certain private individuals may believe this will be" must be declared to be "entirely just and best," "even if it is in some way mistaken" (863e-864a).¹³ The last qualification is necessary because, as we saw, not even the Stranger's city—which he calls the best city possible in practice—will be able to avoid serious inadequacies in its understanding of what is best.

The tripartite psychology of crime presented here by the Stranger is repeated in the immediate sequel with an alteration: in the second version "the striving for

¹³ There exists a controversy in the literature over the exact interpretation of *kan sphalletai ti* at 863a4. I am inclined to read *kan sphalletai ti* as referring to the possibility that the "opinion about what is best" can result in both errors of practical judgment and in mistaken judgments about what the best actually is. Some, following Arthur Adkins, favor the 'good conscience' view, according to which "If a man's reason and desires are not in conflict, *whatever his basic view of life*, he is to be termed *dikaios* [just], provided that his actions are based on reason, not passion or desire." Arthur Adkins, *Merit and Responsibility*, 307-308; italics in original. Others—including O'Brien, Saunders, and Mackenzie—deny that Plato could have admitted that just men could err substantially about the nature of the best, and favor the interpretation of *kan sphalletai ti* as referring only to errors regarding circumstances that lead to harm. See Saunders, "The Socratic Paradoxes in Plato's *Laws*," 430-432; Mackenzie, *Plato on Punishment*, 247-49; Michael O'Brien, "Plato and the 'Good Conscience': *Laws* 863E5-864B7," 84-85. On my interpretation of Book IX (and of the *Laws* as a whole), Plato understands there to be a very great gulf—a difference in kind rather than degree—between the just man in the political sense and the truly virtuous man who possesses knowledge about the human good.

expectations and true opinion concerning what is best” is said to be the third cause of crimes, replacing ignorance in the first list (864a-b). The Stranger does not explain this substitution and appears to regard the repetition as identical to the original (cf. 864b). The striving for true opinion presupposes ignorance about what is best. As we have learned, the opinions of the legislators about the noble, the just, and the good things come to sight as confused upon critical inspection, and thus may be correctly said to partake of ignorance to some extent. Someone like Socrates, who discovers for himself the ignorance of the legislators and reasonably decides to strive for the truth, even if this should bring him to engage in pursuits that run contrary to the settled assumptions of the city about what is best, would be considered unjust on the basis of the Stranger’s criminological scheme. Yet the Stranger leaves it undetermined whether such ignorance should be construed by the city as a light and forgivable fault or as an arrogant claim to wisdom that ought to be punished. In this way, the Stranger’s criminological scheme allows for someone like Socrates to be judged more leniently than the latter had been judged by the Athenians.

In the remainder of Book IX, the Stranger completes his penal legislation for the future city. Like the penal laws concerning capital crimes, the rest of the penal code resembles in part traditional Greek penal practice, although it also contains some humanizing innovations, following the Stranger’s enlightened view of punishment. Throughout, the penal code uses the language of medicine to compare crime to disease, although it also retains the conventional distinction between voluntary and involuntary injustice and applies harsher punishments to the former than to the latter. The principle of

moral responsibility is affirmed so emphatically that at one point in the code even beasts and inanimate objects are said to be capable of committing murder (873e-874a).

Perhaps the best means of true moral reform and the most innovative feature of the Stranger's penal code is not the penalties themselves, but the critique of the legislators' opinions about punishment that constitutes the heart of Book IX. The only arrangement that could bring actual penal practice to conform entirely to the noblest conception of punishment would be the institution of the philosopher judge, who would educate each criminal individually, insofar as he is curable, instead of punishing him. Sweeping political reform in full conformity with the noblest conception of punishment would require a Rhadamanthus in addition to a Minos (cf. 624b-625a with 630c). Yet the institution of the philosopher king is not taken seriously in the *Republic*, and it is not even mentioned in the more politically moderate *Laws*.

Conclusion

The present study began by acknowledging the existence of a moral crisis in modern criminal justice. That crisis manifests itself most clearly as the concern for the moral treatment of criminals in light of the clashing aims that our penal institutions are thought to serve. It is evident, on the one hand, that political society must have some right to protect its individual members against crime, and to this end it appears to be justified in pursuing a deterrence policy in dealing with offenders. Yet, while society's right to self-defense cannot be denied, it is also apparent that the modern, state-sponsored, deterrent penal regime has a tendency to treat criminals callously, and occasionally even to sacrifice their interests to the security of the majority. Thus criminals, in turn, would seem to have a moral claim on society to be treated with some minimum of respect, and even to be cared for and corrected as fellow citizens and human beings who, despite their better selves, have fallen into misfortune. But this custodial and rehabilitative approach to punishment can tend as well to deny the moral responsibility or autonomy of offenders, and thus to degrade them, at least from the modern point of view. Finally, society seems to have a right—nay, a duty—to give to each what he deserves, which seems to require holding each criminal accountable for his choices, as an independent and responsible being, by making his penalty reflect the crime. Yet retribution understood in this way also jars with our conviction that justice must ultimately benefit and not harm.

In practice, penal institutions can appear to pursue all of these different ends at once, which is troubling to many conscientious observers—death penalty abolitionists,

restorative justice advocates, retributivists, etc.—who perceive the clashing of these ends. One reaction to this moral crisis has been to construe the clashing of these aims as the outcome of competing and mutually exclusive principles influencing modern penal practice at the same time. On a theoretical level, such a view would make it necessary for us to choose the most acceptable principle among the alternatives, and then to try to construct a normative theory of punishment based upon it. This approach has not been very successful, however, because it cannot quite defeat the strong claims of the competing principles which are ruled out. For example, purely deterrent or purely retributivist theories of punishment have found relatively few supporters.

Another reaction to the crisis has been to deny that these different aims of punishment are really clashing. Instead, as the “mixed” retributivism theorists suggest, they might be thought of as “answering different questions” about how to justify punishment. This approach has obvious appeal inasmuch as it tries to give the claim of each side its due. And yet by assuming that an unproblematic synthesis of these different aims can be achieved, this approach ignores the genuine tensions that in fact do exist among them.

I argued that in order to better understand what is at stake in each one of these claims, so as to gain a firmer grip on the problem, we needed to take a historical view of the development of modern punishment theory, as I have done in this study. In Hobbes, we saw what is entailed for our understanding of crime and punishment if we take our sights solely by the fragility of civil order and by what is required to preserve it. From this perspective, the demands we place on justice are drastically reduced. Justice must

above all secure civil peace, and it must do so by threatening and carrying out certain and sufficiently terrifying punishments, even if this implies that civil society must be acknowledged to have no moral or legal obligation to its most dangerous criminals and be allowed to destroy these as enemies at war.

If the preservation of civil society means the preservation of *all*, then justice only requires of each individual what is already in his own best interest. On this view, crime can only be regarded as contrary to one's own advantage, and thus as a failure of practical deliberation on the part of the criminal, rather than as the manifestation of a base or vicious character. If both the criminal and the law-abiding citizen pursue what appears to them to be their own good, and if the value of that good is subjective to each, then the criminal cannot be blamed as base or punished with a view to his morally depraved character. Retributive punishment therefore becomes irrational and unjust and deterrence becomes the sole legitimate aim of punishment. And while rehabilitation can still be an appropriate goal of punishment on this view, its meaning is dramatically altered from the traditional one of moral improvement. A Hobbesian policy of criminal reform would amount to making past offenders more clear-sighted about what is necessary for their own comfort and security and better able to control their irrational passions. It would not urge offenders to repent of their base selfishness. Hobbes is, therefore, the father of modern deterrence theories of punishment, and he shows much more clearly than contemporary deterrence theorists that the principle which makes deterrence the *sole* aim of punishment—i.e., the right to self-preservation (not general utility)—effectively rules out the other two alternatives—retribution and rehabilitation—or, at the very least, it

dramatically transforms their meaning. At the same time, Hobbes had a clear enough grasp of moral psychology to see that an unadorned system of pure deterrence would be unpalatable to ordinary moral sensibilities. He thus saw the need for some kind of admixture of deterrent and non-deterrent elements which, at the same time, would not restrict the deterrent capacity of the whole system, although, as we saw, Hobbes himself was unable to accomplish this.

Kant accepted Hobbes's basic analysis of political life as organized with a view to avoiding the ever looming threat of the violent state of nature, but saw that the purely deterrence-based approach to punishment which was implied as a consequence of that analysis was morally repugnant. Kant saw—as many contemporary critics of modern criminal justice now see—that a strictly deterrent penal system would threaten the human dignity of the individual offender. But unlike contemporary restorative justice advocates, Kant did not believe that the rehabilitative approach to crime could be an adequate solution to this problem. Although rehabilitative punishment is intended to promote the well-being of the criminal himself, and thus appears to treat him as an end in itself, it is at the same time not incompatible with paternalistic practices (e.g., those identified by critics of the penitentiary) that relegate the criminal to the role of an irresponsible dependent of the state.

Kant claimed that human dignity and freedom demand that each criminal be held responsible for his actions by being punished in accordance with his desert—no more and no less. But unlike contemporary retributivist theorists, Kant was sensitive enough to the moral psychology of righteous indignation to see that retribution could not be explained

merely as a way of protecting criminals from disproportionate government coercion, or as a way of equalizing the burdens of civil society, or as an emphatic expression of society's notions of right and wrong. All of these alternatives seem to fall short of the deepest meaning of righteous indignation (by which we are inspired to seek retribution in the first place) an insight which Kant ultimately owes to the classical tradition, in its Biblical expression, rather than to his Enlightenment rationalism. According to Kant, the moral consciousness demands that the wicked be made to suffer in proportion with their moral turpitude, just as the virtuous deserve to prosper in proportion with their moral goodness.

But as our study of the complexities and tensions in Kant's account of retribution showed, retributive punishment as he understood it could not be sustained by modern political principles, despite his best attempts. Retributive justice takes into consideration the criminal's moral character in meting out his penalty, but since Hobbes, modern politics had strictly separated justice from morality. Kant respected this separation when he distinguished between the juridical and moral aspects of action, defining the former as its external form—which may be coerced—and the latter as its inner motive—which may not. This separation between justice and morality fatally undermines Kant's claim that retributive punishment must be part of the universal law which every citizen authorizes as a co-legislator in the civil union. Thus, although he saw the moral defects of the modern penal system, and although he tried to remedy these defects by attempting to justify retribution on modern principles, Kant ultimately did not succeed.

Despite his failure, critical reflection on Kant's penal thought is of great value for us in our effort to grasp the present moral crisis of modern criminal justice. Kant's

attentiveness to moral psychology allowed him to see that retribution in its deepest meaning—i.e., the proportional distribution of happiness in accordance with moral goodness—cannot be reconciled with modern principles, unless retribution itself is diluted and assimilated to deterrence (e.g., „fair distribution of burdens,’ ‘protection from disproportionate punishments’) and thus deprived of its deepest moral significance for us. Our study of Kant ultimately showed us that the modern approach cannot be the way to the sought-after solution. We thus found ourselves in a position to begin afresh by looking outside of the modern tradition for guidance. I then argued in favor of following in a more thoroughgoing way those of Kant’s instincts which led him to look to the classical tradition for insights regarding the moral treatment of criminals. I thus suggested that we should try to follow the classical approach to the question of punishment from its very beginning.

We focused on the discussion of the Athenian stranger’s penal code in Plato’s *Laws* as the text which provides the clearest expression of the classical approach to the question of punishment. The most distinctive feature of that approach, as we saw, was its attentiveness to, and cautious respect for, the ends of political life as those ends are understood by the non-philosophic but serious men who live that life. Unlike Hobbes, the Stranger did not narrow the scope of politics to preservation of civil peace, and thus was not compelled to limit the purpose of punishment to crime prevention. On the contrary, since the end that politics sets for itself is virtue as a whole, the Stranger is able to define just punishment as, in the best case, the moral reform of criminals by the city, which is both beneficial to the criminal who is reformed and noble to the city which reforms him.

At the same time, the Stranger also recognized the ineradicable power of the passion of spiritedness in political life, which constitutes an important theme of Plato's political philosophy as a whole. Spiritedness motivates political men to become passionately concerned with retribution, and the Stranger chooses to accommodate this concern while gently tempering it through various institutional reforms. One such reform is to infuse the authoritative penal prelude, and the penal code itself, with the paradigm of medicine, according to which crime is compared to a disease of the soul and punishment to the cure which heals it. The city to which the Stranger gives shape through his legislation will, somewhat confusedly, punish its criminals with a zeal for retribution—when it does punish them—and yet it will also pity them as suffering from a self-inflicted evil and thus will be inclined to punish them less savagely and with greater concern for their moral rehabilitation. In this way, the Stranger elaborates a penal system which is in practice both effective and as humane as possible, although this comes at the unavoidable cost of creating apparent confusion in the aims of the penal law.

Perhaps the Stranger's most important contribution to humanizing his penal system is his critical examination and discussion of the opinions of conventional legislators. This searching examination points to the highly unconventional and momentous observation that the fundamental cause of crime is the deep-seated doubt about the goodness of justice that pervades the whole city and even the city's founders. Through the cross-examination of his non-philosophic fellow-legislator Kleinias, the Stranger shows that human beings both hope that the just life is the best and happiest life, and at the same time doubt that this is so and suspect that they may stand to profit

through injustice. This doubt leads men to vacillate about the choiceworthiness of justice throughout their whole lives and to commit crime on those occasions when they are most tempted by the spoils of crime and least threatened by punishment.

This deepest meaning of the Stranger's discussion of penal laws is not made explicit but only hinted at by his puzzling and provocative statements in the course of the dialogue. Yet it is this profound teaching that is also the most humanizing part of the whole discussion of penal laws. It reveals crime in a completely new light that explains why it is an ineradicable feature of political life. Insofar as we find this explanation persuasive—and in order to find it persuasive we must put it to the test of experience—we are reconciled with the necessity of crime, and are no longer tempted to attribute it to willful human wickedness. This goes a very long way towards purging the angry soul of its righteous indignation. In addition, this new way of looking at crime shows why penal laws are necessary, even though, as the Stranger puts it, they are in a certain sense shameful. This, in turn, can help relieve the moral anguish that we might otherwise feel when we punish our fellow citizens while wishing for some nobler, more humane form of crime control that would obviate the need for punishment altogether.

Works Cited

Primary Sources

- Aristotle. *Politics*. 1984. Trans. Carnes Lord. Chicago: University of Chicago Press.
- _____. *Nicomachean Ethics*. 2002. Trans. Joe Sachs. Annapolis, MD: Focus Publishing Company.
- Beaumont, Gustave de and Alexis de Tocqueville. 1979. *On the Penitentiary System in the United States and its Application in France*. Trans. Francis Lieber. Carbondale, IL: Southern Illinois University Press.
- Beccaria, Cesare. 1996. *Of Crimes and Punishments*. Trans. Jane Grigson. New York: Marsilio Publishers.
- Benn, S. I. 1958. "An Approach to the Problems of Punishment." *Philosophy* 33 (127): 325-341.
- Bentham, Jeremy. 2009. *The Rationale of Punishment*. Ed. James McHugh. Amherst, NY: Prometheus Books.
- Boonin, David. 2008. *The Problem of Punishment*. New York: Cambridge University Press.
- Bradford, William. 1793. *An Enquiry How Far the Punishment of Death Is Necessary In Pennsylvania*. Philadelphia: Dobson.
- Braithwaite, John. 2002. *Restorative Justice and Responsive Regulation*. New York: Oxford University Press.
- Brubaker, Stanley C. 1988. "Can Liberals Punish?" *The American Political Science Review* 82 (3): 821-836.
- ____ and Ronald J. Terchek. 1989. "Punishing Liberals or Rehabilitating Liberalism." *American Political Science Review* 83 (4): 1309-1316.
- Christie, Nils. 1977. "Conflicts as Property." *British Journal of Criminology* 17: 1-26.
- Clark, Ramsey. 1970. *Crime in America*. New York: Simon and Schuster.
- Corlett, J. Angelo. 2001. "Making Sense of Retributivism." *Philosophy* 76 (295): 77-110.
- Dagger, Richard. 1993. "Playing Fair with Punishment." *Ethics* 103 (3): 473-488.

- Feinberg, Joel. 1970. *Doing and Deserving*. Princeton: Princeton University Press.
- Foucault, Michel. 1977. *Discipline and Punish*. Trans. Alan Sheridan. New York: Vintage Books.
- Furman v. Georgia*. 1972. 408 U.S. 238.
- Gregg v. Georgia*. 1976. 428 U.S. 153.
- Hahn, Paul. 1998. *Emerging Criminal Justice: Three Pillars for a Proactive Justice System*. Thousand Oaks, CA: Sage Publishing Company.
- Hart, H. L. A. 1968. "Prolegomenon to the Principles of Punishment." *Punishment and Responsibility*. Oxford: Clarendon Press.
- Hirsch, Adam J. 1992. *The Rise of the Penitentiary: Prisons and Punishment in Early America*. New Haven, CT: Yale University Press.
- Hobbes, Thomas. 1994. *Elements of Law*. Ed. J. C. A. Gaskin. Oxford: Oxford University Press.
- _____. 1994. *Leviathan*. Ed. Edwin Curley. Indianapolis: Hackett.
- _____. 1998. *De Cive*. Trans. Richard Tuck and Michael Silverthorne. Cambridge: Cambridge University Press.
- The Holy Bible: King James Version*. 1991. New York: Ivy Books.
- Jefferson, Thomas. 1943. *The Complete Jefferson*. Ed., Saul Padover. New York: Duell, Sloan and Pierce.
- Kant, Immanuel. 1956. *Critique of Practical Reason*. Trans. Lewis White Beck. New York: Liberal Arts Press.
- _____. 1964. *Groundwork for the Metaphysics of Morals*. Trans. H. J. Paton. New York: Harper & Row.
- _____. 1965. *Critique of Pure Reason*. Trans. Norman Kemp Smith. New York: St. Martin's Press.
- _____. 1980. *Lectures on Ethics*. Trans. Louis Infield. Indianapolis: Hackett Publishing Company.
- _____. 1987. *Critique of Judgment*. Trans. Werner S. Pluhar. Indianapolis: Hackett Publishing Company.
- _____. 1996. *Metaphysics of Morals*. Trans. Mary Gregor. Cambridge: Cambridge University Press.
- _____. 1996. *Religion within the Bounds of Reason Alone*. Trans. George di Giovanni. *Religion and*

- Rational Theology*. New York: Cambridge University Press.
- Lewis, C. S. 1970. "The Humanitarian Theory of Punishment." *God in the Dock: Essays on Theology and Ethics*. Grand Rapids, MI: W. B. Eerdmans Publishing Company.
- Locke, John. 1982. *Second Treatise of Government*. Ed. Richard Cox. Wheeling, IL: Harlan Davidson.
- Machiavelli, Niccolo. 1985. *The Prince*. Trans. Harvey Mansfield. Chicago: University of Chicago Press.
- Menninger, Karl. 1966. *The Crime of Punishment*. New York: Viking Press.
- Mill, John Stuart. 1978. *On Liberty*. Indianapolis: Hackett Publishing Company.
- _____. 2001. "April 1866 Speech on Capital Punishment." *Utilitarianism*. 2nd ed. Ed., George Sher. Indianapolis: Hackett Publishing Company.
- Montesquieu, Charles Louis de Secondat, Baron de. 1989. *Spirit of the Laws*. Trans. Anne Cohler. New York: Cambridge University Press.
- Morris, Herbert. 1968. "Persons and Punishment." *The Monist* 52: 477-501.
- Murphy, Jeffrie G. 1973. "Marxism and Retribution." *Philosophy and Public Affairs* 2 (3): 217-243.
- Pangle, Thomas L., ed. 1987. *The Roots of Political Philosophy: Ten Forgotten Socratic Dialogues*. Ithaca, NY: Cornell University Press.
- Pestritto, Ronald J. 2000. *Founding the Criminal Law: Punishment and Political Thought in the Origins of America*. DeKalb, IL: Northern Illinois University Press.
- Plato. 1921. *Laws of Plato*. Ed. England, E. B. 2 vols. Manchester: University of Manchester.
- _____. 1926. *Laws*. Ed. Jeffrey Henderson. Cambridge, MA: Harvard University Press.
- _____. 1980. *The Laws of Plato*. Trans. Thomas L. Pangle. Chicago: University of Chicago Press.
- _____. 1980. *The Republic of Plato*. Trans. Allan Bloom. New York: Basic Books.
- _____. 1984. *Apology of Socrates*. In West, *Four Texts on Socrates*, 63-97.
- _____. 1984. *Euthyphro*. In West, *Four Texts on Socrates*, 41-61.
- _____. 1987. *Alcibiades I*. Trans. Carnes Lord. In Pangle, *The Roots of Political Philosophy*, 175-221.

- _____. 1987. *Lesser Hippias*. Trans. James Leake. In Pangle, *The Roots of Political Philosophy*, 281-299.
- _____. 1987. *Minos*. Trans. Thomas L. Pangle. In Pangle, *The Roots of Political Philosophy*, 53-66.
- _____. 1987. *Lovers*. Trans. James Leake. In Pangle, *The Roots of Political Philosophy*, 80-90.
- _____. 1998. *Gorgias*. Trans. James H. Nichols Jr. Ithaca, NY: Cornell University Press.
- Primoratz, Igor. 1989. "Punishment as Language." *Philosophy* 64 (248): 187-205.
- Rawls, John. 1955. "Two Concepts of Rules." *The Philosophical Review* 64 (1): 3-32.
- _____. *A Theory of Justice*. Revised ed. Cambridge, Massachusetts: Harvard University Press, 1999.
- Rothman, David J. 1971. *The Discovery of the Asylum*. Toronto: Little, Brown, and Company.
- Rousseau, Jean-Jacques. 1964. *The First and Second Discourses*. Trans. Roger D. Masters and Judith R. Masters. New York: St. Martin's Press.
- _____. 1978. *On the Social Contract*. Trans. by Roger D. Masters and Judith R. Masters. New York: St. Martin's Press.
- _____. 1990. *Collected Writings of Rousseau*. Ed. Roger Masters and Christopher Kelly. 11 vols. Hanover, NH: University Press of New England.
- Rush, Benjamin. 1954. *A Plan for the Punishment of Crime: Two Essays*. Ed. Negley Teeters. Philadelphia: Pennsylvania Prison Society.
- West, Thomas G. and Grace Starry West, trans. 1984. *Four Texts on Socrates*. Ithaca, NY: Cornell University Press.

Secondary Sources

- Adkins, Arthur. 1960. *Merit and Responsibility*. Oxford: Clarendon Press.
- Brown, Keith. C., ed. 1965. *Hobbes Studies*. Oxford: Basil Blackwell.
- Bruell, Christopher. 1999. *On the Socratic Education*. Oxford: Rowman & Littlefield.
- Byrd, B. Sharon. 1989. "Kant's Theory of Punishment: Deterrence in its Threat, Retribution in

- its Execution.” *Law and Philosophy* 8 (2): 151-200.
- Byrne, Peter. 2007. *Kant on God*. Burlington, VT: Ashgate Publishing Company.
- Cattaneo, Mario A. 1965. “Hobbes’s Theory of Punishment.” In Brown, *Hobbes Studies*.
- Fackenheim, Emil. 1996. “Kant’s Philosophy of Religion.” *The God Within: Kant, Schelling, and Historicity*. Toronto: University of Toronto Press.
- Fleischacker, Samuel. 1992. “Kant’s Theory of Punishment.” In Williams, *Essays on Kant’s Political Philosophy*.
- Galston, William A. 1975. *Kant and the Problem of History*. Chicago: University of Chicago Press.
- Gauthier, David. 1969. *The Logic of Leviathan*. Oxford: Clarendon Press.
- Herbert, Gary. 1995. “Immanuel Kant: Punishment and the Political Preconditions of Moral Existence.” *Interpretation* 23 (1): 61-76.
- Hill, Thomas E., Jr. 1999. “Kant on Wrongdoing, Desert, and Punishment.” *Law and Philosophy* 18 (4): 407-441.
- Johnston, David. 1986. *The Rhetoric of “Leviathan.”* Princeton: Princeton University Press.
- Mackenzie, Mary. 1981. *Plato on Punishment*. Berkeley: University of California Press.
- Macpherson, C. B. 1962. *The Political Theory of Possessive Individualism: Hobbes to Locke*. Oxford: Clarendon Press.
- Martinich, Aloysius. 2005. *Hobbes*. New York: Routledge.
- Murphy, Jeffrie G. 1987. “Does Kant have a theory of punishment?” *Columbia Law Review* 87 (3): 509-532.
- Norrie, Alan. 1984. “Thomas Hobbes and the Philosophy of Punishment.” *Law and Philosophy* 3 (2): 299-320.
- O’Brien, Michael. 1957. “Plato and the ‘Good Conscience’: Laws 863E5-864B7.” *Transactions and Proceedings of the American Philological Association* 88: 81-87.
- Pangle, Lorraine. 2009. “Moral and Criminal Responsibility in Plato’s *Laws*.” *The American Political Science Review* 103 (3): 456-473.
- Pangle, Thomas L. 1976. “The Political Psychology of Religion in Plato’s *Laws*.” *American Political Science Review* 70 (4): 1059-77.

- Saunders, Trevor. 1968. "The Socratic Paradoxes in Plato's *Laws*: A Commentary on 859c-864b." *Hermes* 96 (3): 421-434.
- Scheid, Don E. 1983. "Kant's Retributivism." *Ethics* 93 (2): 262-282.
- Shell, Susan M. 1980. *The Rights of Reason: A Study of Kant's Philosophy and Politics*. Toronto: University of Toronto Press.
- _____. 1997. "Kant on Punishment." *Kantian Review* 1: 116-124.
- Strauss, Leo. 1950. *Natural Right and History*. Chicago: The University of Chicago Press.
- _____. 1975. *The Argument and Action of Plato's "Laws."* Chicago: University of Chicago Press.
- Taylor, A. E. 1965. "The Ethical Doctrine of Hobbes." In Brown, *Hobbes Studies*. Originally published in 1938.
- Tuck, Richard. 1979. *Natural Rights Theories*. New York: Cambridge University Press.
- Tunick, Mark. 1996. "Is Kant a Retributivist?" *History of Political Thought* 17 (1): 60-78.
- Warrender, Howard. 1957. *The Political Philosophy of Hobbes: His Theory of Obligation*. Oxford: Oxford University Press.
- Weinrib, Ernest. 1992. "Law as Idea of Reason." In Williams, *Essays on Kant's Political Philosophy*.
- White Beck, Lewis. *Commentary on Kant's Critique of Practical Reason* (Chicago: University of Chicago Press, 1960)
- Williams, Howard, ed. 1992. *Essays on Kant's Political Philosophy*. Cardiff: University of Wales Press.

Vita

Arthur Shuster attended Newtonbrook Secondary School, in Toronto, Ontario. In 1999, he enrolled at the University of Toronto, from which he graduated with a Bachelor of Science in Human Biology, with honors. In the Fall of 2004, Arthur enrolled in the Ph.D. program in the Department of Government at the University of Texas at Austin.

Permanent Address: 300 Bloor Street East #501, Toronto, Ontario M4W 3Y2.

This manuscript was typed by the author.