THE MIRROR OF JUSTICE:  
A PLEA FOR MERCY IN CONTEMPORARY LIBERAL THEORY

A Dissertation

Submitted to the Graduate School  
of the University of Notre Dame  
in Partial Fulfillment of the Requirements  
for the Degree of  

Doctor of Philosophy

by

Daniel Patrick Moloney, B.A., M.A.

_______________________________  
David K. O’Connor, Director

Graduate Program in Philosophy  
Notre Dame, Indiana  
April 2004
THE MIRROR OF JUSTICE:
A PLEA FOR MERCY IN CONTEMPORARY LIBERAL POLITICS

Abstract

by

Daniel P. Moloney

As usually defined, the concepts of justice and mercy seem incompatible—if justice is the strict application of the law, and mercy lenient deviation from it, then mercy is unjust and justice is merciless. Perhaps for this reason, liberal political philosophers have mostly neglected the topic of mercy, despite its traditional role in contributing to political stability. The present study suggests that one can integrate mercy into liberal political philosophy only after significant departures from the usual accounts.

St. Anselm of Canterbury gave the classic formulation of the paradoxes of justice and mercy in his *Proslogion*, and in his later works he solved them. Anselm claims that justice and mercy should be defined in terms of right order or rectitude. Justice is the desire to effect and preserve rectitude, while mercy is the attempt at restoring another to justice so defined. Anselm claims it is both more stable and more humane to *persuade* people to desire right order for its own sake, rather than to *coerce* the people to uphold the political order or to bribe them to pursue it out of momentary advantage.
Mercy, understood in an Anselmian fashion as ordered toward rectitude, is a stabilizing policy when exercised with prudence. The contractarian theories of justice promoted by John Rawls and others are unstable because they cannot earn the support of people who reject autonomy as the organizing principle of their lives and politics. Moreover, it is difficult to develop an adequate account of criminal punishment that is consistent with contractarian liberalism. A view that puts stability and mercy at the center of its accounts of governing and punishing can serve as a more stable foundation for liberal politics than can contemporary views.

The theory developed here sees citizens in democracy as leaders responsible for promoting their ideas of social order. When advocating their ideas, they ought to persuade those who disagree rather than to coerce them, for this promote stability. Likewise with criminals: although, punishment is justified as a defense of societal order against the criminal’s attack on it, it is better to persuade the criminal to obey the law voluntarily and for principled reasons.
## CONTENTS

ABBREVIATIONS ........................................................................................................ v

PREFACE ..................................................................................................................... vi

PART ONE: IS A STABLE, HUMANE LIBERALISM POSSIBLE?

INTRODUCTION TO PART ONE .................................................................................. 1

CHAPTER 1: RAWLSIAN TOLERANCE & ITS ENEMIES

1.1 Tolerance and Humanity in Rawlsian Justice ....................................................... 7
1.2 Political Liberalism's Drive for Stability ............................................................... 12
1.3 Is Political Liberalism Rortyan Pragmatism? ....................................................... 15
1.4 Overlapping Consensus and the Personally Normative ....................................... 18
1.5 Rawlsian Justice and Its Enemies ......................................................................... 21

CHAPTER 2: THE IMPERATIVE OF STABILITY & THE LIBERALISM OF FEAR

2.1 The Imperative of Stability .................................................................................. 26
2.2 How Stable Is Political Liberalism? ..................................................................... 31
2.3 Can Good Christians Be Good Liberals? ............................................................. 33
2.4 The Humane Basis of the Liberalism of Fear ....................................................... 41

CHAPTER 3: THE PUBLIC, THE PRIVATE, & THE BURDENS OF JUDGEMENT

3.1 The Public and the Private .................................................................................. 53
3.2 Locke: The Care of Souls Is a Private Affair ....................................................... 55
3.3 Rawls on the Original Position and Public Reason ............................................. 58
3.4 Publicity and Privacy in the Liberalism of Fear .................................................. 70
3.5 Conclusion ......................................................................................................... 78

CHAPTER 4: LIBERAL AUTONOMY OR HUMANE PUNISHMENT?

4.1 Liberalism & Punishment .................................................................................... 80
4.2 Rawls on Punishment ........................................................................................ 82
4.3 A Short History of Punishment Theories ............................................................ 84
4.4 Does Rawls Have a Theory of Punishment? ....................................................... 94
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5</td>
<td>Punishment in Murphy’s Rawlsian Liberalism</td>
<td>99</td>
</tr>
<tr>
<td>4.6</td>
<td>Deterrence and the Liberal State</td>
<td>107</td>
</tr>
<tr>
<td>4.7</td>
<td>Murphy’s Dilemma: Liberalism or Punishment?</td>
<td>109</td>
</tr>
<tr>
<td>4.8</td>
<td>Punishment and Moral Education</td>
<td>110</td>
</tr>
<tr>
<td>5.1</td>
<td>Anselm on God’s Mercy and Justice: An Overview</td>
<td>128</td>
</tr>
<tr>
<td>5.2</td>
<td>Tools of Natural Theology: Anselm on the Attributes of a Simple God</td>
<td>132</td>
</tr>
<tr>
<td>5.3</td>
<td>Concepts as Mental Words</td>
<td>134</td>
</tr>
<tr>
<td>5.4</td>
<td>Mental Words and the Limits of Understanding</td>
<td>136</td>
</tr>
<tr>
<td>5.5</td>
<td>Anselm’s Purification Procedure</td>
<td>139</td>
</tr>
<tr>
<td>5.6</td>
<td>Pure Perfections</td>
<td>144</td>
</tr>
<tr>
<td>5.7</td>
<td>Are Pure Perfections Really Any Help?</td>
<td>152</td>
</tr>
<tr>
<td>6.1</td>
<td>The Development of Justice and Mercy in Anselm’s Speculative Works</td>
<td>159</td>
</tr>
<tr>
<td>6.2</td>
<td>Anselm on Justice and Mercy in the Oratio ad Sanctum Prolepse</td>
<td>160</td>
</tr>
<tr>
<td>6.3</td>
<td>Justice in the Monologion</td>
<td>163</td>
</tr>
<tr>
<td>6.4</td>
<td>Goodness, Mercy, and Justice: Navigating the Proslogion</td>
<td>169</td>
</tr>
<tr>
<td>6.5</td>
<td>Justice in the ‘Three Treatises’</td>
<td>176</td>
</tr>
<tr>
<td>6.6</td>
<td>Truth, Rectitude, and Justice in De Veritate</td>
<td>179</td>
</tr>
<tr>
<td>6.7</td>
<td>Freedom Is the Ability To Will Justly: De Libertate Arbitrii</td>
<td>186</td>
</tr>
<tr>
<td>6.8</td>
<td>Freedom and the Right Ordering of Creation: De Casu Diaboli</td>
<td>190</td>
</tr>
<tr>
<td>6.9</td>
<td>Justice as the Limit of Mercy: Honor and Rectitude in the Cur Deus Homo</td>
<td>198</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>211</td>
</tr>
<tr>
<td>7.2</td>
<td>Anselm’s Monastic Approach to Punishment</td>
<td>213</td>
</tr>
<tr>
<td>7.3</td>
<td>Anselm’s First Theory of Punishment and Governance: Strict Justice and the Demands of Order</td>
<td>216</td>
</tr>
<tr>
<td>7.4</td>
<td>Anselm’s Second Theory of Punishment: Restoring Order by Merciful Solicitude</td>
<td>228</td>
</tr>
<tr>
<td>7.5</td>
<td>Human Mercy, Political Roles, and the Responsibility for Right Order</td>
<td>247</td>
</tr>
</tbody>
</table>
PART THREE: TOWARDS AN ANSELMIAN LIBERALISM

INTRODUCTION TO PART THREE ................................................................. 256

CHAPTER 8: THE LIBERALISM OF MERCY

8.1 Introduction ........................................................................................................ 258
8.2 Anselmian Principles of Political Justice ......................................................... 260
8.3 Can There Be Contemporary Applications of Monastic Principles? ............. 268
8.4 The Rights and Responsibilities of Citizens .................................................... 273

CHAPTER 9: PUNISHING FOR STABILITY

9.1 Introduction ........................................................................................................ 281
9.2 The Paradoxes of Mercy and Justice ............................................................... 282
9.3 Anselmian and Other Approaches to Punishment ........................................... 287
9.4 Is Anselmian Punishment a Rehabilitative Theory? ....................................... 291
9.5 Is Anselmian Punishment a Moral Education Theory? ................................... 297
9.6 Is Anselmian Punishment a Restorative Justice Theory? ............................... 313
9.7 Indeterminate Sentencing and Social Defense ............................................... 323

CONCLUSION ................................................................................................. 330

BIBLIOGRAPHY .............................................................................................. 333
ABBREVIATIONS

WORKS OF JOHN RAWLS

LP  The Law of Peoples. Harvard University Press, 1999

WORKS OF ST. ANSELM OF CANTERBURY

Aep  Epistolae Anselmi (Anselm’s Correspondence)
CDH  Cur Deus Homo (Why the God-Man)
DC  De Concordia Praescientiae et Praedestinationis et Gratiae Dei cum Libero Arbitrio (On the Concord of God’s Foreknowledge and Predestination and Grace with Free Choice of the Will)
DCD  De Casu Diaboli (On the Fall of the Devil)
DG  De Grammatico (On ‘Grammarian’)
DIV  De Incarnatione Verbi (On the Incarnation of the Word)
DLA  De Libertate Arbitrii (On Free Choice of the Will)
DV  De Veritate (On Truth)
M  Monologion
P  Proslogion

The critical edition of Anselm’s works is the six volume edition by F. S. Schmitt, O.S.B., Sancti Anselmi Cantuariensis Archiepiscopus Opera Omnia (Thomas Nelson and Sons, 1946-61). The letters of St. Anselm are in volumes 3-5. For research purposes, I used the CD-ROM version produced by the InteLex Corporation, which omits volume 6, Schmitt’s index, in favor of an electronic search engine.

TRANSLATIONS OF ANSELM’S WORKS

In the pages that follow, I will be concerned with addressing the seeming incompatibility of justice and mercy. This incompatibility can be stated quite simply: if justice is strict application of the law and mercy lenient deviation from that strict application, then mercy appears to be unjust and justice to be merciless. This is how the two concepts get played out in our common understanding; the tension between them creates the dramatic tension in Measure for Measure, for instance. Consider the plea that Isabella makes for her brother’s life to Angelo who has the power to spare it:

“No ceremony that to great ones longs,  
Not the king’s crown, nor the deputed sword,  
The marshal’s truncheon, nor the judge’s robe  
Become them with one half so good a grace  
As mercy does.  
If he had been as you, and you as he,  
You would have slipped like him, but he like you  
Would not have been so stern” (2.2.60-66).

To which Angelo responds:

“It is the law, not I, condemn your brother.  
Were he my kinsman, brother, or my son,  
It should be thus with him: he must die tomorrow” (2.2.82-84).

Throughout the play, justice is seen as stern, impersonal, and strict, but necessary for a virtuous society, while mercy is portrayed as kind and gentle and in the self-interest of all, but leading to a moral laxity that might corrupt society.
Philosophers and legal theorists also tend to assume that mercy is necessarily in tension with justice. Jeffrie G. Murphy sketches the following objection against allowing mercy into criminal sentencing.¹

If mercy requires a tempering of justice, then there is a sense in which mercy may require a departure from justice. (Temperings are tamperings.) Thus, to be merciful is perhaps to be unjust. But it is a vice, not a virtue, to manifest injustice. Thus mercy must be, not a virtue, but a vice—a product of morally dangerous sentimentality. This is particularly obvious in the case of a sentencing judge. We (society) hire this individual to enforce the rule of law under which we live. We think of this as ‘doing justice,’ and the doing of this is surely his sworn obligation. What business does he have, then, ignoring his obligations to justice while he pursues some private, idiosyncratic, and not publicly accountable virtue of love or compassion?

Murphy here articulates the reasoning that underwrites the strict sentencing guidelines that law and order politicians often impose when significant portions of the judiciary acquire a reputation for excessive leniency.² If judicial mercy leads to widespread crime, there is a predictable demand to restrict the discretion a judge has to mitigate the punishment demanded by law. On this view, mercy seems to be a vice opposed to justice.

On the other hand, Western literature abounds with examples where individuals are treated cruelly in the name of justice and the common good, making mercy seem desirable. This creates a dilemma.

Murphy describes two paradoxes about the relationship of justice and mercy:

First, the two concepts are defined as opposites, and so any system that tries to deploy


both would seem straightforwardly inconsistent. Second, the exercise of mercy seems to be inescapably arbitrary, and if justice is closely associated with fairness, arbitrariness seems unjust. Interestingly, Murphy appeals to St. Anselm of Canterbury as providing “the clearest statements of the paradoxes…on mercy” in chapters 9 and 11 of the

*Proslogion*.

Anselm’s first paradox is how God, who is always just, can grant eternal life to someone who in justice merits eternal death. Murphy notes that, “although Anselm’s specific worry is about the divine nature (are the divine attributes of perfect justice and perfect compassion coherently ascribable to the same being?), he raises a general worry about the concepts of justice and mercy themselves.”

If mercy is simply an aspect of justice—say, the prudent application of universal laws to particular cases—then it is reducible to justice and is not an independent virtue. Mercy, which we tend to believe is something gratuitous, would then really be an obligation, and “all the talk about gifts, acts of grace, supererogation, and compassion becomes quite beside the point.” But if mercy requires or permits that justice sometimes not be applied, then it counsels injustice.

Murphy thinks that this paradox is a permanent feature of criminal justice, and thus “there is no room for mercy as an autonomous virtue with which [a judge’s] justice should be tempered.” Since in criminal justice, the judge is obligated to uphold justice, he is prohibited from exercising any mitigating discretion. The upholding of justice is

---


4 Ibid., p. 169.

5 Ibid.

6 Ibid., p. 174.
thus a perfect duty for a judge in criminal law. In private law, however, justice is a right but not an obligation. In *A Merchant of Venice*, Shylock has a right to exact a pound of flesh from Antonio, but he is free to show mercy by waiving his right to payment.

Murphy thinks that in the private law paradigm, “there is no contradiction, paradox, or even tension” between mercy and justice. Indeed, Murphy suggests that Anselm could solve his theological puzzle about the divine attributes by moving from the criminal law paradigm to the private law paradigm, so that God’s mercy “may be viewed as His deciding out of love or compassion, to waive certain rights that He has—not to violate certain obligations that He has.” In which case, “Anselm’s first paradox disappears.”

But even in the private law paradigm, there does seem to be what Murphy calls “a kind of ‘equal protection’ paradox.”

If God (or any other rational being) shows mercy, then the mercy must not be arbitrary or capricious but must rather rest on some good reason—some morally relevant feature of the situation that made the mercy seem appropriate….But once a reason always a reason. And does not the Principle of Sufficient Reason require that if I, as a rational being, showed mercy to Jones because of characteristic C, then it is presumably required of me (rationally required, not just morally required) that I show comparable mercy to C-bearing Smith?

If this logic holds, Murphy suggests, then mercy would not be obligatory; “but if I slip and show it even once, then I am rationally required to show it to all relevantly similar persons.” In practice, that might deter people from showing mercy at all!

---

7 Ibid., p. 176.
8 Ibid., p. 177.
9 Ibid., p. 181.
Murphy thinks he can solve this paradox for human justice, if not for divine justice, by proposing that mercy is what Kant called an imperfect duty, which “admits of wide latitude in the time and manner of its fulfillment.”¹⁰ We all want a society in which people in general are merciful, argues Murphy, but we don’t want mercy to be obligatory, for the reasons given. So we want people to be merciful sometimes and not all the time. The poor and weak are more likely to want mercy than the rich and powerful, while the rich and powerful are more interested in justice (conceived as social stability and regular enforcement of the rules). The rich make a deal with the poor, so that the poor are shown mercy some of the time, in exchange for which the poor cooperate and comply with the rules of society, creating stability. This doesn’t make sense when applied to divine mercy, but perhaps our moral views about mercy are simply the result of just such a tacit bargain advancing the self-interests of both the rich and powerful and the weak and poor.

This puzzle about justice and mercy will occupy Part Two of this dissertation, when we look at how Anselm solved the paradoxes he articulated. It was necessary to touch on this puzzle first, though, because of a certain stereotype about mercy. For complicated historical reasons, mercy has come to have religious connotations, and we have forgotten that politics is the term’s original home.¹¹ Mercy is a virtue only for one who has some sort of power with relation to another—either power over another, or the power to help another. Perhaps mercy remains a topic for legal philosophers considering

¹⁰ Ibid., p. 183.

¹¹ An Internet search on ‘mercy’ turns up references to religious orders, papal homilies, and hospitals almost exclusively. Interestingly, the term ‘merciless’ seems to have retained some of its political connotations, especially for sportswriters, journalists describing military massacres, and for comic book villains.
the criminal justice system because the practice of criminal punishment is one place in which our society clearly exercises coercive power over its citizens, creating a situation in which the question of mercy still presents itself.

There are, of course, many areas outside of the criminal justice system in which power is exercised. For instance, it has even become a cliché in our information age to assert that power depends on knowledge, or to put it terms of the bumper sticker equation, Knowledge Is Power. It is much less common to observe the deep connection between epistemology, the study of knowledge, and politics, the study of power. But the connection is there for all to see. Not only do recent political theories—those of Rawls, MacIntyre, and Rorty, e.g.—evince this connection, but so do older ones. Kant thought that justice was intimately tied to rationality, conceived in his particular, critical way. Adam Smith developed an economics that was compatible with a Humean skepticism about our ability to know beyond our experiences. Montaigne and Locke both saw, in their own ways, that an awareness of our ignorance about the highest matters could be used to force politics towards a tolerant peace based on low, but solid and evident, truths. In a less skeptical vein, Plato’s Republic, perhaps the first work in political philosophy, sees political justice depending on knowledge of the Forms. The quasi-platonic empire known as Christendom held that government should be based on the privileged access to the Truth given by God to His one, true Church. Islamic and Jewish theocracies made similar assumptions.

Given this general connection between theorizing about knowledge and theorizing about power, it should come as no surprise to notice that there are some close
relationships between particular political regimes and particular epistemological views. The separation of church and state characteristic of modern liberalism, for example, required a critical number of the right sort of influential people to embrace the premise that a practical knowledge of God’s will in this life either was not possible or was too difficult to institutionalize at a scale commensurate with governing structures as large as the nation-state. A certain sort of skepticism became sufficiently accepted that the leaders of sixteenth- and seventeenth-century Europe decided to structure their governments around the search for peace, since that seemed easy to define and possible to achieve, rather than the search for God or Truth, which they came to think was not.\textsuperscript{12}

The resulting cultural pluralism, which so flourishes in our day, has put pressure on philosophers in our day to rethink this fundamental option for peace over truth. This pressure has come from a number of different sources. The dogged refusal of the world to achieve peace has for cynics made “peace on earth” an empty eschatological phrase, on the order of “when hell freezes over.” The criticisms by those grouped under the rubric, first of Marxism, later of postmodernism, have shown how the commitments of liberal ‘peace’ often function as subtle theoretical and epistemological truth claims that preserve and promote a not uncontroversial set of power relations. Finally, the clamoring of those from other cultures for the opportunity to live their cultures to the fullest has created headaches for those whose job it is to promote harmony where liberalism’s version of harmony is not always valued.

In the face of such criticisms, some philosophers have wondered whether liberals ought not just assert that liberalism is good and true, and abandon the skeptical and compromise-based epistemologies that tended to give it the upper hand over more dogmatic theological-political belief systems. Others have thought that liberals should simply admit that liberalism is not about some abstract justice, but is merely a way of controlling warring factions, so that at bottom it is about balancing powers and redirecting political energies to minimize intolerable or cruel outcomes, and so will never rise above the politics of the modus vivendi. Some have proposed alternative epistemologies that will lead to different political structures, grouped around the goods of community and the cultivation of virtue. Of all these voices, only John Rawls and his followers seem to have preserved what we suggested were liberalism’s primordial commitments to peace at the level of politics and neutrality at the level of epistemology. Indeed, when Rawls looked in the mirror some time in the late 1970s and saw that his theory of justice-as-fairness could lead to an unstable politics as a result of its departure from epistemological neutrality, rather than abandoning these commitments, he preferred to spend the last twenty years of his life trying to fix his highly regarded theory.

As we will see in the first three chapters below, the doctrine of tolerance—long seen as the key liberal virtue that secures peace through an epistemological neutrality between different worldviews—leads to just the sort of political instability Rawls seeks to guard against. Yet it is upon tolerance (also known in his later writings as reasonableness or a sense of justice) that Rawls bases his whole system. For Rawls and for many liberals, tolerance is often secured by skepticism about inegalitarian claims to know important truths, most often religious claims about the good. This is because liberal tolerance is a
horizontal virtue, regulating the attitudes and behaviors between those assumed to be equals. But if I assert that I know the truth and you don’t, or if I condemn you for acting *evilly* (rather than unfairly, say), then I am making an inegalitarian claim: In the first case, I claim that you are not my equal in our understanding of the most important realities. In the second, I claim that there is a standard of good and evil that applies to you whether or not you agree, that you have fallen short of it, and (at least implicitly) that I have not fallen short in the same way. Since claims about truth and falsehood, good and evil, if significant enough will have inegalitarian political consequences, it has been a strategy of liberalism since its inception to control such claims, either by urging their removal from public life, or by launching skeptical attacks against them, or by persecuting those who persist in asserting them in public. This last strategy has been perhaps a more recent development, dating to some Supreme Court cases in the 1940s, but it arguably was implicit in the logic of tolerance all along that certain sorts of inegalitarian behaviors would have to be intolerable.

But as long as truth claims and assessments of good and evil persist, so will their inegalitarian political consequences. At the end of chapter three below, we will see how tolerance’s inability to allow certain sorts of truth claims will require a regime based on this doctrine to deploy its state power in ways that are likely to lead to widespread instability. And at the end of chapter four, we will see that in the deliberations of those philosophers who attend to the central political institution of punishment, where inegalitarian judgments about morality and the good are necessary and inescapable, the values of fairness and tolerance enter hardly at all. Much has been written about the first weakness in the doctrine of tolerance, little about the second. But both point to a
weakness in the doctrine that, we will see, is a result of its assumption of a controversial sort of equality.

Part One thus argues that Rawls’s arguments for autonomy and his arguments for stability pull apart, that stability is the more basic and important characteristic, and that the drive to promote autonomy can sometimes have a destabilizing effect. The argument of Part One shows the need for a way to stabilize a liberal regime that doesn’t try to impose a view of autonomy.

Given the above, it seems that liberals need to foster a virtue that will allow them to handle certain sorts of inegalitarian behavior while preserving their commitments to peace and stability amidst cultural and intellectual pluralism. That is, they need a vertical virtue (or set of virtues) to govern the appropriate behaviors and attitudes among those who are not equals, or who at least refuse to lay aside their claims to superior knowledge of the true and good. Historically, this virtue was called mercy.

In Part Two, we will examine what is perhaps the most theoretically sophisticated doctrine of mercy, the extremely influential view developed by Anselm of Canterbury. Indeed, although Anselm is not usually considered an important political thinker, contemporary philosophers who think about punishment still return to his ideas about guilt, culpability, redemption, atonement, and also mercy.\footnote{In addition to Murphy, “Mercy and Legal Justice,” see also Nigel Walker, “The Quiddity of Mercy” Philosophy v. 70 (1995) pp. 27-37; William Kneale, The Responsibility of Criminals (Oxford University Press, 1967).}

The broadly Platonic tradition of reflection on justice sees human justice as related in some quasi-univocal way (usually by a relation of partaking or participating)
with ideal Justice itself. The broadly Aristotelian tradition of reflection on justice derives principles of justice by abstraction and generalization from common human practices (usually commercial, political, and penal practices), and consequently tends to draw a stark conceptual distinction between divine and human justice. Anselm sits in the broadly Platonic tradition as just described. He assumes that Justice is providential, the justice of the governor, and thus justice is intrinsically related, not to equality, but to order. Since human justice should imitate divine Justice, it too should be conceived in terms of order rather than equality. The paradoxes highlighted in the *Proslogion*, which for Murphy are primarily political, are for Anselm at once conceptual and deeply metaphysical, even theological. How can God be both Supreme Justice and Supreme Mercy, if mercy and justice are in tension, as they seem to be?

Chapters five through seven work together as a set piece, laying out Anselm’s theory of justice and mercy. Chapter five looks at Anselm’s theory of pure perfections, his theory about why we should use our understanding of the attributes of God to control our understanding of the attributes of man. The details of this theory provide the tools, the motivation, and much of the direction for the development of Anselm’s theory of justice and mercy, the topic of chapter six. In that chapter, we will examine the major works of Anselm’s corpus in some detail to show how Anselm first develops and then solves the paradoxes of mercy that Murphy lays out.

Anselm lays out his theory of just governance in the context of explaining why sinners need to be punished and why they need to be punished eternally. In chapter seven, then we will lay out Anselm’s view of justice by looking at his theory of how God ought to punish sinners. We will then examine his argument why it is necessary for a powerful
and good governor to be merciful as well as just. This chapter briefly touches on some recent scholarship by historians trying to piece together Anselm’s political actions as Archbishop of Canterbury. Their results show many similarities with the view of justice and governance developed in Anselm’s speculative works.

In Part Three, we try to develop an Anselmian liberalism, adapting Anselm’s ideas and principles as elaborated in Part Two to the circumstances of modern democracies. Chapter eight begins with a summary of the political principles found in Anselm. Anselm’s political principles all apply to rulers, and were conceived and articulated within a feudal society. In order to translate these principles for modern democracies, we introduce *The Principle of Monastic Liberalism*:

In democracies, all citizens can be thought to have the role of a ruler or governor with real but very circumscribed powers, and thus real though diffuse responsibilities.

This principle instructs us to consider each citizen as a ruler, hearkening back to a view of the democratic citizen that emphasizes his responsibilities for governing more than emphasizing his rights. When democratic citizenship is conceived in this way, each citizen finds himself responsible for promoting the stability and right ordering of society. Each citizen, therefore, will be obligated to employ mercy when doing so would increase the stability and right order of the regime. It is for this reason that we call the liberalism sketched here ‘the liberalism of mercy.’

Chapter nine shows how Anselmian liberalism applies to the practice of criminal punishment. In chapter four we argued that it was a serious, even fatal weakness of Rawlsian liberalism that it did not have a sound theory of punishment. Because Anselm’s political principles were designed with the problem of punishment in mind, Anselmian liberalism is not subject to this objection. Indeed, the chapter goes on to argue, Anselmian liberalism includes a theory of punishment that has significant advantages over the five major theories of punishment: theories retribution, deterrence, psychological rehabilitation, moral education, and restorative justice.

If nothing else, the dissertation shows that Anselm’s views of justice are relevant to our time, despite the claims of many critics. At best, the argument here shows a new, more stable way of looking at politics and criminal justice, one in which the virtue of mercy plays a significant humanizing and stabilizing role.

---

PART ONE

INTRODUCTION TO PART ONE:

IS A STABLE, HUMANE LIBERALISM POSSIBLE?

In *A Theory of Justice*, Rawls forcefully placed questions about the right, the good, and man’s sense of his own self-worth at the heart of moral and political philosophy. He proposed that morality should first of all concern itself with justice, and justice should first of all concern itself with tolerance and liberty of conscience, so long as philosophers regard man’s sense of his own self-worth to be a key moral value and protecting that sense of self-worth to be a key political value. Rawls would require that society’s institutions guarantee each person the widest latitude of personal behavior consistent with reason and public order. His liberalism tries to eliminate most of the occasions in which the law’s impersonal character would conflict with the most personal commitments of those it governs. If Rawls had his way, tolerance, liberty of conscience, and liberty of opportunity would be written into the constitution and counted among the very principles upon which the just society is based. Indeed, Rawls places liberty of conscience as the first principle of a just regime.

This is relevant for our discussion of mercy in two ways. First, when we tolerate someone whose views or behavior we find disagreeable or simply wrong, rather than using coercion to correct the person, we exercise something like mercy towards that
person. So any discussion of the political virtue of mercy ought to examine the role that
tolerance plays in modern society, and Rawls’s work is central to the academic
conversation about this. Rawls’s work is also relevant to discussions of mercy because
Rawls’s theory is able to reconcile justice and tolerance—indeed, his theory of justice is
built around tolerance—and so it would seem that any attempt to reconcile mercy and
justice ought to begin by examining how theorists did the trick with tolerance.

In the first chapter, we will see how Rawls is able to reconcile these values,
though at a certain cost. The Rawlsian regime would be inhumane to a certain class of
citizen, whom Rawls thinks it must treat inhumanely if it is to treat the rest of the citizens
in a way that would preserve or increase their sense of self-worth. Rawls is sensitive to
this concern, and in his later work he revises his theory to restrict the size of this class. He
does so, not so much out of the moral concern that his theory allows for inhumane
treatment of some citizens, as out of the political concern that if a large enough class of
citizens rejects the very principles of a regime, that regime will not be stable.

In chapter two, we will take up this insight that a stable regime treats its citizens
humanely,¹ and from this angle look at another view of liberal politics. Known as the
‘liberalism of fear,’ this family of theories claims that a person’s instinctive horror at
cruelty—of being treated cruelly oneself or seeing others cruelly treated—is the moral

¹ I use the word ‘humane’ here and throughout in the sense of the Latin
humanities, which sometimes translates the Greek philanthropia or love of mankind. Its
opposite is ‘inhumane.’ Martha Nussbaum appeals to this sense of humanitas in the title
of her book Cultivating Humanity (Harvard University Press, 1998), as does Alain
Finkelkraut in the title of his book In the Name of Humanity (Columbia University Press,
2000).
reaction most central to liberalism.\(^2\) This sort of liberal is uncertain, bordering on skeptical, about any grandiose plans to organize a just society. On the other hand, she is very certain that cruelty in any form is just plain wrong, and that a good society is organized to minimize the chances that any person will be treated cruelly. Since torture and other sorts of cruelty are most widespread when done by powerful agents of the state, this sort of liberalism is especially concerned with limiting concentrations of power.

Judith Shklar, who coined this term, sees Montesquieu as the father of this sort of liberalism, and Montaigne as its paradigmatic sensitive soul.\(^3\) Richard Rorty tries to claim Rawls for this camp, arguing that in the articles that led up to *Political Liberalism*, as the more personalistic elements of Rawls’s theory came to dominate its more legalistic and moralizing aspects, Rawls’s view came to approach Rorty’s own pragmatic, skeptical, and anti-utopian liberalism that puts few demands on those it governs other than that they get along and let each other pursue their own personal projects.\(^4\) Philip Quinn tries to enlist Christians into a liberalism of fear, “broadly construed,” i.e., without some of the elements that are explicitly skeptical or otherwise opposed to the Christian tradition.


\(^3\) Cf. Shklar, “The Liberalism of Fear.”

Quinn argues that Christians and liberals share a deep aversion to cruelty and an opposition to coercion regarding religious belief, and that even for Christians, these personalistic concerns outweigh the impersonal concerns of maintaining orthodoxy and removing sin and vice from the world.\(^5\)

The second chapter concludes that there might be a liberalism of fear which extends the protections of liberalism to the class of citizens that Rawlsian liberalism is unable to accommodate. Chapter three looks at the thorniest issue for such a liberalism: the distinction, essential for any liberal theory, between the public and the private realm. Rawls bases his version of the public/private distinction on a particular understanding of autonomy, and a key result of the argument of this chapter finds that building a regime around this idea of autonomy would lead to just the sort of instability that Rawls seeks to avoid. An amended version of the liberalism of fear can avoid this problem, but at the cost of assuming that the relation of the state to its citizens (and of the citizens, through the state, to each other) will be characterized by heteronomy, though hopefully a benign heteronomy.

The fourth chapter presents a different reason to replace a liberalism based on autonomy with one based on benign heteronomy, by taking up the topic of punishment. Punishment is obviously a central function of government (indeed, having a monopoly on coercive power is sometimes thought to be government’s defining characteristic), and it is also a first-class example of (not necessarily benign) heteronomy. The deeply heteronomous character of punishment causes deep and obvious problems for the

\(^5\) Cf. Quinn, “Can Good Christians Be Good Liberals?”
contractarian and other autonomy-based liberal theories dominant in the academic discussion. We will show this by investigating two attempts at integrating a theory of punishment into a contractarian liberalism: Rawls’s own half-hearted efforts in *A Theory of Justice*, and Jeffrie Murphy’s more developed arguments on behalf of a Kantian/Rawlsian retributivist theory of punishment. Rawls’s view seems internally incoherent and only partially worked out, while Murphy’s view is terrifically unattractive (as he himself later realized). What’s more, neither theory is available to a Rawlsian political liberal. The fourth chapter concludes by examining some theories of punishment, advanced by contemporary liberals, which argue that ‘paternalistic’ punishment might be the only sort that is an improvement over mere domination.

The various arguments in Part One thus converge on a single conclusion. If liberalism is to be viable, its goal cannot be to abolish heteronomy, but rather must consist in the effort to make it more benign. A regime built around heteronomy of any sort will require that, at least in some respects, citizens do not regard themselves as equals. At this point, though, tolerance can no longer serve as the sovereign virtue—as the term has come to be used in contemporary discussion, tolerance applies between equals. Rawls claims that ‘civic friendship’ characterizes liberal democracies, a phrase that evokes classical views of the friend as another self, one’s equal in the most important respects. On the other hand, ‘paternalism’ is the derogatory term used to describe what I have (perhaps euphemistically) called benign heteronomy. Part One concludes that a stable society needs to be built upon a virtue that is proper to the father, but that can also be extended to the friend and the self. That virtue is mercy. Reintroducing mercy to the discussion, however, reintroduces some of the problems mentioned at the beginning of
how to reconcile mercy and justice. Solving those problems is the task of Part Two.

Applying the solution to the case of liberal politics is the task of Part Three.
CHAPTER ONE

RAWLSIAN TOLERANCE AND ITS ENEMIES

1.1 Tolerance and Humanity in Rawlsian Justice

A social contract theory begins with the individual person and develops an idea of politics that most (ideally all⁶) individuals can come to accept. It follows from this that any social contract theory is likely to place fundamental human concerns at the center of public life, so that politics never departs too far from what most people will accept. If well designed, social contracts are thus founded upon a solid sense of humanity. Of course, the hypothetical nature of the social contract requires an idealized understanding of the human beings who will agree to the contract. Real live constitutional conventions bring too many special interests to bear, so that the resulting social contract is not guaranteed to reflect the interests of all those who live under it. It is often observed, for instance, that the United States Constitutional Convention of 1787 included only

⁶ Jeremy Waldron’s classic essay, “Theoretical Foundations of Liberalism,” (Philosophical Quarterly, v. 37, 1987, pp. 127-150) presents the “fundamentally liberal thesis” (p. 140) as requiring that all those governed by a social and political order must consent to the order, or else its use of force is illegitimate. If taken strictly, this threatens to make liberalism into a form of utopianism—surely one obdurate Tory needn’t delegitimize a liberal form of government. Waldron’s article canvasses several ways that (he thinks) liberals try to meet this very high standard while avoiding utopianism. I’m not sure liberals actually shoot for this standard, and so I don’t want to be committed to defending the exegesis behind Waldron’s thesis.
landholding white men, several of whom had slaves. The resulting constitution allowed
slavery and restricted the franchise to landholding men. Those who make this point
usually suggest that if the convention had been open to a more diverse set, some of the
special interests would not have had their interests written into the final document. This
is one reason why social contract theorists do not rely on the results of actual
constitutional conventions, but imagine their contracts being devised in more ideal
circumstances.

Since the human being who ‘agrees’ to such a hypothetical social contract is of an
abstract and ideal sort, there is room for some debate about which features of real people
get included in the ideal person and which get left out. Different choices about these
features will likely lead to different conclusions about the nature of the contract.
Furthermore, different choices about the features of the ideal participants in the contract
will affect how humane and livable are the real institutions that embody the principles of
the ideal contract. If one assumes that the participants in the ideal contract are without
vice or sin, then the institutions that result might be unworkable when animated by
vicious and sinful real people. Jeremy Waldron puts the point well:

7 This account is too simplistic to do anything more than to observe that the
Constitutional Convention fell short of this ideal of disinterestedness. I intend no moral
judgment on the participants in the Convention.

8 Waldron (“Theoretical Foundations of Liberalism,” p. 140 ff.) lists some other
reasons that theorists gravitate toward hypothetical social contracts, chief of which is that
it would be far too demanding to insist that for a proposed political order to be legitimate
all those governed by it must actually agree to submit to it.

9 Ibid., p. 144.
When we move from asking what people actually accept to asking what they would accept under certain conditions, we focus on the reasons that people might have for exercising their will in one way rather than in another. Doing so involves certain dangers for the liberal. Real people do not always act on the reasons we think they might have for acting; the reasonableness of the actors in our hypothesis may not match the reality of men and women in actual life.

If the contract is to govern a pluralistic society, there are some particularly sticky problems for the liberal theorist. If a contract is to be agreeable to people who share widely divergent ethical beliefs, the theorist has to discover some underlying feature or overarching form of all the ethical commitments present or likely to be present in the society. If he fails, some real people will not be able to agree to the social contract, even in principle, and the liberal’s use of coercive power is, by his own standards, illegitimate. The liberal’s commitment to solving these sticky problems is said to be his commitment to neutrality. According to his own principles, if he violates neutrality, he is treating people unfairly and inhumanely.

Rawls designed his original position and his account of “goodness as rationality” in the *Theory of Justice* with this goal of neutrality in mind. Rawls maintained that the

---


12 Rawls thinks that justice as fairness is superior to teleological theories because the latter tend toward “inhumanity and fanaticism.” See *A Theory of Justice, Revised Edition*, §84, pp. 486-7.
good for each person has a single form. He defines a person’s good as “the satisfaction of a rational desire,”¹³ which he explicates in a formal way:¹⁴

\[ A \text{ is a good } X \text{ for } K \text{ (where } K \text{ is some person) if and only if } A \text{ has the properties (to a higher degree than the average or standard } X) \text{ which it is rational for } K \text{ to want in an } X, \text{ given } K’s \text{ circumstances, abilities, and plan of life (his system of aims), provided that } K’s \text{ plan of life, or that part of it relevant in the present instance, is itself rational.} \]

Rawls then suggests that a human good is a good that it is rational for all persons with rational plans of life to want. The purely formal nature of this goodness leads Rawls to promote its “moral neutrality.”¹⁵ When the principles of justice derivable from the morally neutral original position are added to this neutral definition of goodness, Rawls thinks he has arrived at a morally neutral foundation for the idea of moral goodness in a just society. The constraints of maintaining this morally neutral foundation lead Rawls to develop his original position around a relatively thin conception of the self and a “thin theory of human good.”¹⁶ By keeping his foundational assumptions ‘thin’ and ‘neutral,’ Rawls believes that he can justify his principles of justice to any reasonable person,

¹³ Rawls, ibid., §16, pp. 79-80.

¹⁴ Cf. ibid., §61, p. 351. I’ve compressed the three stages of Rawls’s definition into one stage.

¹⁵ Ibid., pp. 354-5. The idea that ‘the good for persons’ might be a morally neutral concept may seem puzzling. For Rawls, something only earns the adjective “moral” if it depends upon, assumes, or makes reference to the principles of justice. Hence, in the passage cited here, he distinguishes between “goodness” and “moral goodness” depending on whether the “moral principles” of justice are involved or not. For an examination of the idea that “the moral” ought to be a specially demarcated sphere within ethics, and a critique, see Bernard Williams, Ethics and the Limits of Philosophy (Harvard University Press, 1985) pp. 1-7, ch. 10.

¹⁶ For “the thin theory of human good,” see TJ, §§15, 22-25,60-68.
where a reasonable person is one who is willing to recognize and respect in others their commitment to a rational plan of life, whatever that might be. In doing so, he respects what is “perhaps the most important primary good” a person’s sense of his or her self-worth.\(^{17}\)

Rawls sets up his social contract as a contract between idealized reasonable persons who embrace the thin theory of the good and respect the good of self-worth. If his assumptions are not neutral among the commitments of the actual people whose social order is governed by this contract, then Rawls has to go back and look for different assumptions that truly are neutral among the commitments of those governed by the contract. If there are some citizens who cannot accept any liberal social contract, if there are likely to be some indigestible anti-liberals in a society, then Rawls (indeed everyone who adheres to the liberal principle of legitimacy) faces the following “hard choice,” to use Waldron’s words:\(^{18}\)

Either he concedes that his conception of political judgment will be appealing only to those who hold their commitments in a certain ‘liberal’ spirit. Or he must look for a form of social order in which not only those with different ideals, but those with different views about the legitimacy of imposing their ideals, can be accommodated.

In the years following the publication of *A Theory of Justice*, critics identified several entire categories of citizens who could not accept the minimal commitments of Rawls’s

\(^{17}\) TJ, §67, p. 386. See also, §82, pp. 477-9.

initial theory. So Rawls went back and restricted his principles even further, resulting in a new political (rather than intellectual or metaphysical) justification of justice as fairness, which he called political liberalism.

1.2 Political Liberalism’s Drive for Stability

Political liberalism places the goal of liberal legitimacy front and center, arguing that unless justice as fairness can win adherents from all corners of a pluralistic society, the exercise of political power by a liberal regime would not be “fully proper.” Rawls develops this program as a response to ‘the problem of stability.’ In A Theory of Justice, which the later Rawls classifies as a work in moral, as opposed to political, philosophy, Rawls introduced the idea of a well-ordered society, in which all the citizens of the liberal regime endorse justice as fairness as a comprehensive moral doctrine, and accept the two principles of justice as rooted in this doctrine. If everyone in society shares the same moral doctrines, Rawls believed, the society would be stable, in that nobody would threaten to undo or fundamentally alter the constitution and basic structures of society in ways that would threaten the fundamental ideas of liberalism. Consequently, the regime of justice as fairness would perdure over time from generation to generation.

---


20 PL, p. 137.

21 This characterization of a well-ordered society is Rawls’s own. Cf. PL xviii.
In a series of articles beginning in 1980 and culminating in the book *Political Liberalism* and the lecture “The Idea of Public Reason Revisited,” Rawls submitted his *moral* idea of a stable, well-ordered society to what he believed to be a devastating *political* critique. The idea of a well-ordered society in which everyone embraces justice as fairness is politically unrealistic, he argues, because it is simply a fact that modern democratic societies are characterized by *reasonable pluralism*. By this he means that their citizens subscribe to a variety of comprehensive doctrines about the good which are incompatible with each other (hence *pluralism*), yet which all respect the fundamental assumptions of democratic society, as well as the autonomy and rights of their fellow citizens (his special sense of *reasonable*). The idea of a well-ordered society in *A Theory of Justice* requires that all members of a democratic society abandon their own reasonable comprehensive doctrine for justice as fairness. Rawls realized that however desirable this might be as a moral ideal, it was quite unrealistic as a matter of practical politics, and therefore this part of the theory would have to be reworked.

In *Political Liberalism*, Rawls stipulates three general facts about the political culture of a democratic society:

1) *The fact of reasonable pluralism.* Reasonable pluralism is a permanent feature of liberal democracies, because it results from the exercise of practical reason by people able to exercise their liberty of conscience within the framework of free institutions. A Catholic or Orthodox Jew, for example, can develop a theoretical justification of liberal institutions by moving back and forth between her experience of living

---


23 PL, pp. 36-38.
in a free society and her experience of living as a Jew or Catholic. The result, a Catholic or Jewish liberalism, would be a reasonable comprehensive doctrine that, because it is rooted in the experience of liberal democratic institutions, is native to modern democracies.

2) *The fact of oppression.* The oppressive use of state power is the only way to maintain unity around one comprehensive doctrine.

3) In a democratic regime, a substantial majority of the politically active citizens must endorse and willingly support the regime or else it descends into religious and class conflict and becomes unstable and unlikely to survive. Rawls does not give this fact a name, but we might call it *the fact of democratic consensus.*

From reasonable pluralism and democratic consensus, Rawls concludes that if it were possible for all reasonable citizens to affirm the same set of narrowly defined practical political conclusions, a democratic regime could ground its politics upon those conclusions in a stable and enduring fashion. Rawls calls this the idea of an *overlapping consensus,* and he says that a (truly) well-ordered society is one where this overlapping consensus is broad and is grounded on the reasonableness of the citizens (in Rawls’s special sense of ‘reasonable’).

The second fact, the fact of oppression, also plays a role in the well-ordered society. Rawls admits that *unreasonable* comprehensive doctrines, those which do not forswear using the coercive powers of the state to their advantage, will always exist. A well-ordered society, though, will “contain them so that they do not undermine the unity and justice of society.”

Rawls is thus prepared to admit that “the oppressive use of state power” must be used to defend liberal institutions from the politically active adherents of what he calls unreasonable doctrines. Rawls believes that democratic consensus will erode when oppressive state power is used too often, and so the stability of the regime of

---

24 PL, pp. xviii-xix. See also, p. 64 n.19, pp. 152-153.
political liberalism depends on there not being a substantial minority of politically active citizens who are unreasonable (in Rawls’s sense).

Faced with Waldron’s ‘hard choice,’ then, Rawls takes the first option, viz., to recognize that political liberalism will not appeal to everyone, but only to those of an already ‘liberal bent.’ This interpretation of *Political Liberalism* is not always recognized, so we need to defend it here briefly.

1.3 Is Political Liberalism Rortyian Pragmatism?

In an article first published in 1988, before the publication of *Political Liberalism* but after Rawls had published many of the articles signaling his political turn, Richard Rorty argued that Rawls’s later articles revealed showed him to have been a pragmatist all along. On Rorty’s reconstruction, the doctrine in *A Theory of Justice* was not a normative theory of justice, but a “historic-sociological description of the way we live now” (the ‘we’ refers to “we liberals”). Parts of *A Theory of Justice* do lend themselves to this interpretation. Rawls recognized that his theory would not be acceptable to ‘the intolerant,’ by which he meant those who were not willing to endorse the liberty of conscience, who refused to recognize and respect in others their commitment to a rational

25 Rorty, “Priority of Democracy to Philosophy,” op.cit. Cf. p. 185: “Rawls’s writings subsequent to *A Theory of Justice* have helped us realize that we were misinterpreting his book, that we had overemphasized the Kantian and underemphasized the Hegelian and Deweyan elements.”

26 Ibid. p. 185. See also, p. 189: “Rawls is not attempting a transcendental deduction of American liberalism or supplying philosophical foundations for democratic institutions, but simply trying to systematize the principles and intuitions typical of American liberals.”
plan of life if it differed too much from their own. Rawls names Thomas Aquinas and the Protestant Reformers as advocates of intolerance in this sense, and claims that their views are not compatible with the principles derivable from the original position. From this it might seem that the principles of justice as fairness are not meant to appeal to everyone, but only to a pre-selected group of liberal-leaning citizens.

Yet, in a well-ordered society, all plans of life will make reference to the principles of justice. So Rawls clearly does not think that the views of Aquinas or the Reformers would find expression in rational plans of life in the just society. Indeed, Rawls is rather vague about how, within a well-ordered society, an intolerant sect can come to exist at all. His treatment of how to deal with them has more the flavor of a thought experiment regarding the limits of equal liberty than as a serious treatment of real people.

Rorty’s interpretation relies heavily on another section of A Theory of Justice, in which Rawls describes why the pursuit of happiness cannot serve as an overarching,

---

27 TJ §35, pp. 190-194.

28 Ibid., §34, pp. 189-90.

29 Cf. TJ, §69, p. 398: “Now a well-ordered society is also regulated by its public conception of justice. This fact implies that its members have a strong and normally effective desire to act as the principles of justice require.” Rawls spends all of chapter VIII arguing that justice as fairness is likely to meet this standard.

30 He begins the meat of his discussion as follows: “Suppose that, in some way or another, an intolerant sect comes to exist within a well-ordered society accepting the two principles of justice. How are the citizens of this society to act in regard to it?”
“dominant end” that provides an objective ground for all rational plans of life.\textsuperscript{31} After describing Ignatius of Loyola and Aquinas as examples of those who advocating loving and serving God as man’s dominant end,\textsuperscript{32} Rawls argues that “it strikes us as irrational, or more likely as mad,” to pursue one dominant end. “Human good is heterogeneous because the aims of the self are heterogeneous.”\textsuperscript{33} Rorty seizes upon this description of Loyola and Aquinas as mad to argue that this shows Rawls to be addressing his argument only to those who are already part of the community of “we liberal democrats,” who find in Loyola and Aquinas such a fundamental challenge to “our” values and way of life that their arguments should not even be engaged.\textsuperscript{34} If Rorty is right, then Rawls is not doing universal normative philosophy, but something closer to social commentary on the values of a subset of the populace in the constitutional democracies of today.

It seems, however, that Rorty’s reading overlooks the broader context of this passage. Rawls’s rejection of the notion of a dominant end is actually in service of his argument that justice as fairness can be a universal moral norm which can be arrived at through a careful application of rational choice theory, in reflective equilibrium with our


\textsuperscript{32} TJ, p. 485.

\textsuperscript{33} Both quotes are from TJ, §83, p. 486.

\textsuperscript{34} Rorty, “The Priority of Democracy to Philosophy,”p. 187-8: “We heirs of the Enlightenment think of enemies of liberal democracy like Nietzsche or Loyola as, to use Rawls's word, ‘mad.’ We do so because there is no way to see them as fellow citizens of our constitutional democracy, people whose life plans might, given ingenuity and good will, be fitted in with those of other citizens…. They are crazy because the limits of sanity are set by what we can take seriously. This, in turn, is determined by our upbringing, our historical situation.”
considered moral intuitions and certain psychological and scientific facts. Rawls argues, against Aquinas and Loyola, that perfectionist teleological theories—which, as teleological, are committed to the idea that there is one dominant end—are psychologically unrealistic. In the very next section, Rawls develops a parallel argument that utilitarianism fails because of its teleological structure.\textsuperscript{35} As a result of these arguments, Rawls can conclude that teleological theories fail,\textsuperscript{36} and that the only sensible moral view places the right before the good, as justice as fairness does. So it seems that Rawls’s own description is right—\textit{A Theory of Justice} does present a normative comprehensive moral doctrine.\textsuperscript{37}

1.4 Overlapping Consensus and the Personally Normative

Michael Sandel argues that before the book \textit{Political Liberalism} was published, it did seem as though the “political” Rawls was heading in the pragmatic direction Rorty indicated (and welcomed). However, Sandel concludes, “In \textit{Political Liberalism}, Rawls pulls back from this purely pragmatic account.”\textsuperscript{38} Sandel notes that even in \textit{Political

\begin{itemize}
\item \textsuperscript{35} TJ, §84, p. 486: “I shall understand hedonism…as trying to carry through the dominant-end conception of deliberation.”
\item \textsuperscript{36} TJ, §84, p. 490: “The structure of teleological doctrines is radically misconceived: from the start they relate the right and the good in the wrong way.”
\item \textsuperscript{37} This explains the Christian language and overtones of the famous last line: “Purity of heart, if one could attain it, would be to see clearly and to act with grace and self-command from this point of view [of the original position]” (TJ, §87, p. 514). See Matt. 5:8.
\item \textsuperscript{38} Michael Sandel, \textit{Liberalism and the Limits of Justice, Second Edition}. (Cambridge University Press, 1998) p. 194. Paul Weithman shows that Rorty’s reading of the ‘political’ articles was questionable even before the publication of PL. See Paul
\end{itemize}
Liberalism, Rawls has a normative doctrine, even if only a political one: In an overlapping consensus, every person with a reasonable comprehensive doctrine will eventually come to accept a liberal political conception for moral reasons derived from her comprehensive doctrine. Because citizens are committed to reasoning politically from the two principles, the principles will have normative force. Thus, Sandel concludes, Rawls rejects Rorty’s invitation for a purely pragmatic foundation.

Sandel is right that the overlapping consensus is supposed to provide normative reasons for the citizens to accept the two principles of justice. Yet Rorty is also right that the normative reasons are not universally normative, i.e. binding on everyone no matter what. Here’s how to reconcile these claims: Let us say that a reason R is personally normative for person X if X accepts R as binding no matter what. I might think lying is wrong because it violates the categorical imperative, and violating the categorical imperative is personally normative for me. You might not find the categorical imperative to be coherent, so you do not think that lying is wrong for that reason. That is not to say that you think it is okay to lie: you might have some other reason that is personally normative.


40 J. Judd Owen thinks that in PL Rawls explicitly rejects Rorty’s invitation, though without mentioning Rorty by name. Rawls writes, “It would be fatal to the idea of a political conception [of justice] to see it as skeptical about, or indifferent to, truth” (PL, 150). In Owen’s view, Rawls rejects Rorty’s pragmatism as being founded on antifoundationalism, leading to unnecessary conflict with many foundationalist reasonable comprehensive views (and thus to instability). See J. Judd Owen, Religion and the Demise of Liberal Rationalism. (University of Chicago Press, 2001) pp. 103-107.
normative for you that makes lying wrong (it is opposed to the natural law, for instance).
A reason that is personally normative for me might be universally normative, but it need
not be. Rawls thinks that an overlapping consensus on a particular conclusion C can be
universally normative if all the people affirm C for reasons that are personally normative
for them. If the consensus on C is not normative, then we do not have an overlapping
consensus on C, but a mere modus vivendi or compromise among competing
comprehensive doctrines.

Rorty is certainly right that political liberalism is “antiuniversalist,” and thus not
universally normative. Rawls abandons the claim to be universally normative by adopting
the idea of an overlapping consensus. Sandel is also right that political liberalism is not
“purely pragmatic” as Rorty’s account suggests, but normative, even if only personally
normative for all those reasonable people who affirm the principles in the overlapping
consensus. If Rawls were to adopt a purely pragmatic account, Sandel notes, it would be
tantamount to abandoning his claim that the right is prior to the good, because our ideas
about justice would be up for grabs just as much as our ideas about which comprehensive
document to affirm. And Rawls clearly wants to affirm that our ideas of justice can be
determined independently of our views about the good.

Thus, Sandel’s defense of Rawls against Rorty’s reading is successful, but
limited. Rawls is not a pure pragmatist who provides nothing but “a historic-sociological

42 Sandel, Liberalism and the Limits of Justice, p. 194.
description of the way we live now.” He has a normative conception, one which he believes is widely shared, but which he regards as normative even if not widely shared.

But the normative conception is not universally normative, but only personally normative for those who affirm reasonable comprehensive doctrines. Consequently, Rorty’s description of Rawls’s method as “simply trying to systematize the principles and intuitions typical of American liberals” seems rather apt. Rawls hopes that liberal-leaning citizens will be able to agree on principles of justice while disagreeing over their ideals of the good. He has no such hope for the ‘unreasonable’ citizens.

1.5 Rawlsian Justice and Its Enemies

As in *A Theory of Justice*, the political Rawls is able to rule out a priori the existence of unreasonable and intolerant people in his ideal well-ordered society, but only by grasping the first horn of Waldron’s dilemma. In order to respond to the fact of reasonable pluralism and its consequences for the problem of stability, Rawls takes Waldron’s advice and excludes those whom he thinks are incorrigibly anti-liberal from the community of those who make up the social contract. In doing so, Rawls gives up his pretensions to neutrality and admits that political liberalism will find itself ruling over its enemies.

---


44 Rawls, *Justice as Fairness, A Restatement* (Harvard University Press, 2001) p. 24: “I emphasize that the conception of the person as free and equal is a normative conception: it is given by our moral and political thought and practice, and it is studied by moral and political philosophy and by the philosophy of law.” See also PL p. 18 n20.

enemies (as his tough talk about “containing” the unreasonable suggests).\textsuperscript{46} Furthermore, only the reasonable will find that political liberalism supports their sense of self-respect.\textsuperscript{47}

Earlier we said that by the liberal’s own principles, he treats people unfairly and inhumanely if he violates neutrality. Does Rawls then stand convicted of inhumanity for abandoning neutrality? It depends. Rawls’s own version of the “liberal principle of legitimacy” lets him off the hook, for it requires that a government justify its use of coercive force by appeal to those principles and ideals “which all citizens as free and equal may reasonably be expected to endorse in the light of…their common human reason.”\textsuperscript{48} The words in italics are Rawlsian jargon meant to exclude the intolerant citizens (who do not respect others as free and equal) from the legitimating community, i.e., those whose consent is necessary for the legitimacy of the exercise of political

\textsuperscript{46} Cf. Waldron, “Theoretical Foundations of Liberalism,” p. 146: “If this line is taken, we must abandon any claim about the ‘neutrality’ of liberal politics. The liberal will have to concede that he has a great many more enemies (real enemies—people who will suffer under a liberal dispensation) than he has usually pretended to have.”

\textsuperscript{47} Cf. PL 180: “The social bases of self-respect are explained by the structure and content of just institutions together with features of the public political culture, such as the public recognition and acceptance of the principles of justice.” And again, p. 319: “We argue that self-respect depends upon and is encouraged by certain public features of basic social institutions, how they work together and how people who accept these arrangements are expected to (and normally do) regard and treat one another” (emphasis added). Rawls’s comments about the social bases of self-respect make only minor changes from those in TJ. It is still ‘the most essential of the primary goods,’ (PL, p. 319) and is deeply embedded in the theory of political liberalism. See PL, pp. 81-2, 84, 180, 318-320.

\textsuperscript{48} PL, p. 137.
power. By this standard, the regime of political liberalism uses its coercive power legitimately when it “contains” the unreasonable.

On the other hand, political liberalism does fail to meet the more stringent standard of perfect neutrality. The only way it can protect the primary good of self-worth is by restricting the political activity (and thus offending the sense of self-worth) of those who reject the constraints of liberalism. It is willing to thwart the intolerant in their pursuit of their idea of the good, even threatening to throw them in prison or to use military force against them if they try to advance their agenda politically, in order to uphold and protect a set of values and institutions it makes no effort to justify to them. In this respect, Rorty’s account of Rawls’s political turn is quite correct. In the interests of stability, Rawls is forced to confine the humanitarian core of his liberalism—protecting the social bases of a person’s sense of self-worth—to a mere subset of all those living under it. The Rawlsian political liberal places definite limits on

\[49\] Cf. Marilyn Friedman, “John Rawls and the Political Coercion of Unreasonable People,” in *The Idea of a Political Liberalism*, ed. by Victoria Davion and Clark Wolf (Rowman & Littlefield, 2000) pp. 16-33. Friedman’s analysis of Rawls has been a great influence on this part of the argument. It should be noted, however, that her argument assumes Rawls is not presenting an ideal theory in which citizens are reasonable and rational at the start. If Rawls is presenting such an ideal theory, then I think he assumes that there would not arise large numbers of the unreasonable if the society is well-ordered.

\[50\] For the purposes of this argument, I assume that Rawls and political liberals who follow him would agree that Kennedy was right to order the National Guard to force Governor Wallace to integrate the Alabama schools, since Wallace was trying to defend an unreasonable (because racist) agenda. I also assume that they would accept that the U.S. Supreme Court’s decision to threaten the members of the school board of Santa Fe Independent School District with charges of contempt of court if they continued to allow prayer at high school football games (prayer at a public school event being an unreasonable attempt to advance a private religious belief). Cf. *Santa Fe Independent School District v. Doe*, 120 S. Ct. 2266 (2000).
which persons he or she must treat with humanity, and how far those for whom tolerance is optional can go in their political activity before the demands of justice require that they be contained, with coercive force if necessary.51

Rawls designed his political liberalism to reduce the points of intractable dissent in a liberal regime. He hopes that reducing intractable dissent would make the liberal regime stable, given prevalent societal conditions. At this point it will be helpful to step back from the argument to look at the problem of stability again, in a certain degree of isolation from Rawls’s framework. I think that the problem of stability is more fundamental than liberalism; arguably, liberalism first articulated itself as a solution to the instability of the sixteenth and seventeenth centuries, especially the instability caused by the Wars of Religion and the breakup of Christendom.

I propose to look at liberalism itself through the lens of stability.52 If we do, we will find that stability is closely linked with the qualities of humanity and mercy: The


52 Tomasi (Liberalism Beyond Justice, chapter 1) conceives of political liberalism as a development over ‘ethical’ liberalism because it approaches closer to the liberal principle of legitimacy, explained either in Waldron’s terms or in Rawls’s terms (as described above). I see the advantage of political liberalism in terms of its greater stability, given pluralistic modern societies. Tomasi’s own improvements in political liberalism also increase stability, in no small part because they make it easier for a liberal regime to justify its use of coercive force to all its reasonable citizens. Legitimacy improves stability, because it facilitates consent. But as Rawls points out in his discussion of stability in TJ, “However attractive a conception of justice might be on other grounds
more a regime attends to the needs of its subjects, the more stable it is. Impersonal justice, on this view, can be a principle of instability, since it tries to move people from where they are to where justice demands that they be, whether or not they want to move there or are able to change in the ways such justice requires. I argue that the problem of stability encourages us to put mercy first, before justice. I also argue that the problem of stability creates for political leaders the imperative of stability—the stability of the social order must be a more basic priority for them than the advancement of other values, or else they are negligent as rulers. Together, these arguments uncover an imperative for leaders to put mercy first.

[legitimacy, e.g.], it is seriously defective if...it fails to engender in human beings the requisite desire to act on it” ($69$, p. 398).
CHAPTER TWO

THE IMPERATIVE OF STABILITY & THE LIBERALISM OF FEAR

2.1 The Imperative of Stability

In both *A Theory of Justice* and *Political Liberalism*, Rawls argues that a stable regime is one in which the citizens support it from principled reasons and not simply out of momentary advantage. If the citizens have this principled acceptance of the regime, they will work to sustain its institutions and pass on its ideals to the next generation. Presumably, they will overlook the imperfections in the people who hold the positions of authority, because they accept the rule of law rather than men. Rawls thinks stability is important in part because political society is supposed to be thought of as existing in perpetuity: “it produces and reproduces itself and its institutions and culture over generations and there is no time at which it is expected to wind up its affairs.”\(^1\) It seems reasonable to conclude with Rawls that perpetual societies cannot have their basic structures founded on impermanent alliances.

Rawls points out that by living under a just regime and finding such a life to be attractive and desirable, citizens are likely to move from a mere modus vivendi (in which they support the regime because it is to their advantage to do so) to a stable acceptance of

\(^1\) PL, p. 18.
its principles. Of course, it is also possible to react the other way, to reject the regime and all it stands for. Every regime will have to put up this to some degree. However, there is a limit to how much fundamental dissent a regime can sustain before it becomes unstable and citizens cease to work to sustain it.\(^2\)

The problem of stability is thus not a problem simply within Rawls’s version of liberalism. It is, as Rawls suggests, “fundamental to political philosophy.”\(^3\) In the next part of the chapter we will argue that Rawls is right that stability is fundamental. Indeed, we will argue that it is more fundamental even than Rawls’s principle of liberty. If that is right, and if Rawlsian liberty and autonomy lead to instability, then we have a Rawlsian argument against the principles of justice as fairness.\(^4\)

The problem of stability creates an *imperative of stability*: a government should act to preserve the long-term stability of the regime. Leaders of a regime are responsible for conducting themselves according to principles that garner sufficient support so that

\(^{2}\) George Weigel in *The Final Revolution* (Oxford University Press, 1992) tells the story of the fall of Communism in terms that reinforce Rawls’s point: In his telling, various human rights movements, especially in Poland under the inspiration of Bishop Karol Wojtyla, challenged the principles at the heart of Communist society, proposing a different idea of human nature and human dignity. They gradually won the hearts of the people, who refused to sustain the Communist government, and sapped the government’s will to suppress them.

\(^{3}\) PL, p. xix. In support of Rawls’s claim, we even find a version of the problem in Plato’s *Gorgias* 515e, when Socrates criticizes Pericles for solving the problem of stability domestically by extending the scope of Athenian democracy, but at the cost of corrupting the citizens—another sort of instability—and thus leading to their defeat in the Peloponnesian War.

\(^{4}\) Cf. PL 35-6, 65-6, 167-8.
the regime (i) is able to sustain itself in the face of external and internal attack, and (ii) is able to instill in future generations a respect for the regime and its principles. This imperative follows simply from the responsibility the political leaders incur as leaders. To be a political leader is to be responsible for the continuation of peaceful government of these people, and thus for the stability of the regime.  

As we have seen, instability arises when there is such dissent from the principles animating the regime that those who need to act to sustain the regime are unable or unwilling to do so. When people disagree with the actions of a regime and not its principles, they normally don’t want to alter the form of government itself; they want to change only the people in power. Democracies, for example, remove failed politicians by voting them out of office rather than appointing military dictators or rewriting their constitutions. If the principles of the regime itself come to be identified with the actions of those in power, citizens who are dissatisfied with those actions might opt to change their form of government, but this is not a counterexample to the claim above.  

5 ‘The stability of the regime’ does not mean “the preservation of this government with these people (or this class of people) in power no matter what.” Stability can be a principle that encourages leaders to alter certain features of their regime—as a thought experiment we can easily imagine John Rawls as a leader of a society that was well-ordered in the sense of TJ, who for reasons of stability altered it to approximate the society of PL. This suggests that there are two types of “regime change.” Stable regime change preserves some of the features of the present regime, and alters those features (usually certain of the principles of state action, but sometimes particularly onerous political actors) that seem to be causing instability. Poland’s transition from Communism is a contemporary instance of stable regime change. Unstable regime change is a disruptive and radical break with the features of the previous regime, usually as a result of war or revolution. Poland’s takeover by the Nazis is an instance of unstable regime change, as was the French Revolution.

6 One might wonder whether the imperative of stability is not too strongly stated. Mightn’t it be the responsibility of a leader of an unjust regime to allow the regime to
The imperative of stability, then, requires that political leaders always act from principles likely to garner support sufficient to sustain the regime, and that these principles lead each new generation of citizens to choose to support the regime. Those citizens who accept the principles of the regime will repudiate those leaders who fail to discharge this responsibility. If a regime consistently finds itself attracting unscrupulous leaders, the regime is ipso facto unstable, and the citizens will be forced to replace it with one they think will be more stable.

This imperative means that the leaders of the regime are responsible for at least one aspect of the interior life of its citizens: They are responsible for fostering a certain sort of principled affection for the regime. The need to discharge this responsibility for the interior lives of citizens creates a limit on the way a regime can exercise its power and remain stable. For responsibility is linked to power in two ways:

1. *With power comes responsibility.* If one has the power to perform some action, then whether or not to carry out that action necessarily becomes an object of moral consideration. If man has the ability to destroy the world, then he becomes responsible for deciding whether or not he should destroy the world.\(^7\)

2. *With responsibility comes the need for power.* If one finds oneself with responsibility of some sort, then one is also responsible for acquiring and maintaining the ability to discharge it. So for example, a man who lapse, as many of the Communist leaders did in 1989? That is, aren’t there times when stability ought to give way to justice? I would argue that 1989 is so remarkable precisely because it showed a number of leaders who chose some sort of objective stability when it became clear that it would require a major crackdown to continue to defend the principles of their regime. The peaceful transition of power from one form of government to another preserved stability in a way that a violent defense of the principles of the Communist regimes, or the violent imposition of Western-style liberal democracy, would not have.

discovers he is about to become a father finds himself needing to make enough income to support a family.

We’ll call these the axioms of responsibility. The leader of a regime has the power to rule, and thus the responsibility to rule well. One basic test of a good ruler is whether the people subject to him are better off after he ceases to rule, which entails as a minimum condition that the regime continue to exist after he leaves office. Thus, we have the Imperative of Responsibility: it is one of a leader’s most basic responsibilities to maintain the stability of the regime.

He needs to discharge this responsibility by using a power that is available to him. He lacks the power to compel citizens to assent to the principles of his regime, since the power to command the free assent of another’s will does not exist. That means he must persuade at least some of the citizens to support the principles of the regime. Therefore he must act in ways that his citizens can willingly support. Furthermore, he cannot act in ways that future generations are likely to reject, since that will lead to long-term instability. So the imperative of stability bestows a responsibility on political leaders, and human nature puts certain limits on the powers they can acquire to discharge it.

This creates the moral dynamic that leads to more humane government: A leader meets with significant principled opposition to his actions, and starts to worry that his regime will become unstable. He realizes that by acting from principles likely to appeal to more of his citizens, or to appeal to the most politically active more deeply, he can increase the stability of his regime. He works to discern what principles his people are willing to accept, and negotiates a settlement with them, which leads to what Rawls calls a modus vivendi. Over time, if the principles are acceptable enough, the people come to
accept them for their own sake, and the regime can be said to be stable. Thus, when King John faced significant rebellion from his nobles over his harsh rule, he agreed to sign the Magna Charta, granting them significant rights against the king, in order to save his crown. The principles animating the Magna Charta were enshrined over generations, providing the germ of civil liberties in England.

On this view, a chief goal of political philosophy should be to find principles that are likely to be acceptable to all people at all times. Such principles will have to meet the following criteria, therefore: i) they cannot be based on contingent features of a particular people or time and place, such as the current balance of power between religious groups; ii) they must not assume that a society will be homogeneous. So any political philosophy that seriously addresses the problem of stability will have to find some way of balancing the permanent features of the human condition with the heterogeneity of beliefs and cultures over which a stable regime must rule. Rawls’s political liberalism certainly tries to do this. In what follows, we will scrutinize his views to see in what ways they can be improved upon.

2.2 How Stable Is Political Liberalism?

As we have seen, Rawls’s regime of political liberalism is prepared to treat harshly those ‘unreasonable’ people who are not willing to accept its principles. According to Rawls, the use of coercive power by the state will be legitimate so long as its use is neutral among all reasonable citizens and is in principle justifiable to them.⁸

---

⁸ PL, p. 137.
Several critics of political liberalism argue that one or both of these features makes political liberalism illegitimate or unstable on its own terms.

In *Liberalism Beyond Justice*, John Tomasi argues that⁹

“Political liberalism—even if formally distinct as a justificatory type—*in practice amounts to* the same thing as ethical liberalism….Liberalism has the same transformative and homogenizing implications as ever before. It just now brings about those changes in an indirect and long-term way. To some, political liberalism may even seem to be a kind of fake or fraud: ethical liberalism in stealth mode.”

Many of those to whom political liberalism is directly addressed are committed to liberalism’s political ideals, but have nonliberal ideas in the rest of their life (for example, they don’t adopt the difference principle or the principle of liberty of conscience in their families and churches). Adopting the principles of political liberalism, Tomasi argues persuasively, tends to lead people to develop a whole series of liberal practices and attitudes in their private lives. When those who want to maintain a nonliberal private life realize that their commitment to liberalism in politics will weaken their commitment to their nonliberal private ideals, they may choose their private ideals and weaken their commitment to liberal politics. In this case, the spillover effect of liberal political principles will undermine support for those principles, and lead to instability. Tomasi’s elegant and suggestive book makes several refinements to the positions of political liberalism in order to make them more palatable and fair to a wide array of reasonable people. He argues that political liberals, in the interests of stability, ought to adopt something like the Catholic idea of subsidiarity. That is, they ought to work to make sure

---

that the virtues and values of political liberalism do not overflow into private life, working to maintain them only at the most general levels possible.\textsuperscript{10}

Tomasi’s plans, if workable, would make political liberalism even more humane towards those who meet Rawls’s definition of reasonable. But Tomasi’s amendments to political liberalism do not extend to the unreasonable. He finds it just as necessary to exclude them from the legitimating community.\textsuperscript{11} So even this more modest and refined version of political liberalism will lead to long-term instability in societies with significant populations of unreasonable people. And there are some religious thinkers who argue that many or perhaps most religious citizens will qualify as unreasonable in Rawls’s sense.

2.3 Can Good Christians be Good Liberals?

It is not too strong to say that political liberalism must be made acceptable to devout Christians, or else it fails by its own standards. If liberalism requires that Christians give up their Christianity in order to be liberals, then Christians would presumably not endorse liberalism, leading to widespread instability. One might argue that Christianity is the most likely of the major religions to endorse political liberalism—not only is liberalism an outgrowth of Christian culture, but unlike Judaism and Islam, it has a well-developed traditional account of the proper scope of secular

\textsuperscript{10} Tomasi, \textit{Liberalism Beyond Justice}, p. 127. For the Catholic doctrine of subsidiarity, see Pope Pius XI, \textit{Quadragesimo Annum} (1931) nos. 79-80.

\textsuperscript{11} Cf. Tomasi, \textit{Liberalism Beyond Justice}, p. 21.
authority and sacred authority. If political liberalism requires secularism or encourages it, then it would be unworkable because it would be unstable in the conditions of persistent religiosity that actually obtain in nearly all societies throughout the world. If this worry is right, then an essential part of the broader discussion of liberalism and justice will be the debates within Christianity and other comprehensive doctrines about whether to accept liberalism. The question of stability forces political philosophy to engage certain aspects of philosophy of religion and perhaps of theology.

For this reason, I wish to spend some time discussing a debate about liberalism between two philosophers of religion, Nicholas Wolterstorff and Phillip L. Quinn. In a series of recent articles and books, Wolterstorff argues that religious citizens cannot accept the terms of Rawls’s public reason if it means, as it seems it does, that they have to

---

12 Indeed, on Paul Berman’s account, Sayyid Qutb, the theologian whose ideas are behind the radical Islamism of al Quaeda, thinks that the public-private distinction (especially the separation of church and state) is a principal reason for Muslims to reject Western liberal democracy. See “The Philosopher of Islamic Terror,” NY Times Sunday Magazine, March 23, 2003, p. 24ff. See also Berman’s Terror and Liberalism (W.W. Norton, 2003).

13 Berman ends his New York Times article by urging philosophers and theologians to come to the defense of liberalism: “The followers of Qutb speak, in their wild fashion, of enormous human problems, and they urge one another to death and to murder. But the enemies of these people speak of what? The political leaders speak of United Nations resolutions, of unilateralism, of multilateralism, of weapons inspectors, of coercion and noncoercion. This is no answer to the terrorists. The terrorists speak insanely of deep things. The antiterrorists had better speak sanely of equally deep things. …But who will speak of the sacred and the secular, of the physical world and the spiritual world? Who will defend liberal ideas against the enemies of liberal ideas? Who will defend liberal principles in spite of liberal society's every failure? …Philosophers and religious leaders will have to do this on their own. Are they doing so? …There is something to worry about here, an aspect of the war that liberal society seems to have trouble understanding—one more worry, on top of all the others, and possibly the greatest worry of all.”
conduct politics without making appeals to their deepest religious convictions.\textsuperscript{14} He argues that liberals are mistaken when they assume that it would not violate any person’s religious convictions to debate and act politically on grounds other than religious ones, or to leave religion out of most aspects of public life. The liberal assumes\textsuperscript{15}:

That though religious people may not be in the habit of dividing their life into a religious component and a non-religious component, and though some might be unhappy doing so, nonetheless, their doing so would not be in violation of anybody’s religion. But he’s wrong about this. It’s when we bring into the picture persons for whom it is a matter of religious conviction that they ought to strive for a religiously integrated existence—it’s then, especially, though not only then, that the unfairness of liberalism to religion comes to light.

One might add that there might be some who think that, if they were to say publicly that they have one reason for their political action, and yet have a different private reason that is their real motivation, this would be lying or so close to lying that a delicate conscience would be offended.\textsuperscript{16}


\textsuperscript{15} Wolterstorff, “Why We Should Reject”, pp. 176-7.

\textsuperscript{16} It should be emphasized that Rawlsian public reason prohibits this sort of insincerity. Rawls goes so far as to call it hypocritical when “citizens talk before one another one way and vote another,” (PL p. 215). Rawls insists that all citizens have a duty of civility, which amounts to the proper exercise of their sense of justice in public debate. To be insincere about one’s motives in public discussions about fundamental issues would be to violate this duty, Rawls thinks.
Quinn, in a wide-ranging and insightful discussion of Wolterstorff’s arguments, first strengthens Wolterstorff’s criticism before responding to it.\textsuperscript{17} He seizes on the example of those who “strive for a religiously integrated existence” as the nerve of Wolterstorff’s argument, and focuses his examination there. Quinn points out that Wolterstorff’s arguments aren’t applicable to all Christians, since some Christians seem to be “at home in the differentiated social order to which modernity has given birth.”\textsuperscript{18} Appealing to “sociological grand theory in the tradition of Max Weber and Talcott Parsons,” Quinn observes that modern societies are composed of several differentiated spheres, “such as the economy, politics, religion and so forth,” which are “relatively autonomous and more or less self-contained.” These separated social spheres are what make the separation of politics and religion “a realistic social possibility; such a separation is not feasible in a traditional society that is religiously homogeneous and tightly integrated by its shared religion,” such as Amish societies, or presumably, the Kirias Joel community in New York, or Saudi Arabia. Thus, modernity itself makes it difficult to maintain a religiously integrated existence, according to Quinn’s analysis. Modern Christians, those who have made peace with the differentiated social order, do not have the same struggles as those Christians who reject this social order. So Wolterstorff’s complaints about liberalism only apply to those non-modern Christians, and not to all Christians.

\textsuperscript{17} Philip L. Quinn, “Can Good Christians Be Good Liberals?”; see also “Religious Citizens within the Limits of Public Reason,” \textit{The Modern Schoolman} 78 (2001) pp. 105-124.

\textsuperscript{18} Quinn, “Can Good Christians Be Good Liberals?” p. 7.
There are some people who are not religious but who nonetheless seek a more integrated life. These people, who include “secular Millian liberals and Marxist socialists,” have comprehensive doctrines that “make it a matter of conviction for them that they ought to strive for lives integrated around those doctrines.” Quinn names all these people, whether religious or not, integralists.\(^{19}\) Quinn extends Wolterstorff’s arguments to this group, and suggests that Rawlsian liberalism is “unfair” to integralists in general if it is unfair to integralist Christians. Quinn thinks that Wolterstorff’s arguments might show that political liberalism places a heavier burden on integralists than others, and that as a consequence “people of their sort are bound to decline in numbers and influence in a liberal democratic society in which the Rawlsian ideal of public reason takes hold and becomes more fully realized over time.” But this special burden on integralists, and the increased possibility of their extinction, is not thereby unfair to them, claims Quinn. Fairness on this view is the opposite of arbitrariness; because the burdens imposed by political liberalism can plausibly considered “social necessities,” they would not be arbitrary, but fair and just.\(^{20}\)

Even the relatively capacious culture and institutions of liberal democracy are bound to prove uncongenial to some valuable forms of life, and integralism of various stripes may be among them. …[So] there remains no good reason to suppose that integralists have been unfairly treated merely because their forms of life fail to endure and gain adherents, should such a failure occur in the history of a liberal democracy. Nor is there good reason to suppose that they have been unfairly treated merely because they bear burdens others do not.

\(^{19}\) Ibid., pp. 6-7.

\(^{20}\) Ibid., p. 9.
But even if there is a way to save liberalism from charges of being unfair, there are still the problems of stability that arise when integralist citizens notice that the burdens of political liberalism fall disproportionately on their shoulders. Quinn is sensitive to this worry, and so tries to find a stronger argument that could persuade integralists (or at least, religious integralists) to support the institutions of a liberal polity. He is trying to argue that even religious integralists can support liberalism from within their comprehensive doctrines, thereby including them as potential participants in an overlapping consensus.

His argument has two branches to it, based on a distinction between two sorts of liberalisms: the liberalism of rights as articulated by Locke, Mill, and Rawls; and the liberalism of fear as articulated by Montaigne, Montesquieu, and Shklar. Quinn concludes that integralist Christians can endorse the liberal polity from the theoretical perspective of the liberalism of fear, and that they should endorse (at least some\(^2\)) liberal rights as protective of great goods that they enjoy and that they ought to wish others enjoy.

\(^2\)Quinn’s argument for liberalism of rights comes in two stages: Integralist Christians \textit{can} support the liberalism of rights, because having liberal freedoms is a great good, and there are no necessary reasons that counterbalance these great goods. And they \textit{ought} to support the liberalism of rights, because it supports the great good of searching for religious truth. I think Quinn’s second argument supports only the narrow conclusion that Christians ought to endorse the right to free exercise of religion, and any rights of speech, assembly, etc. that follow from it. It does not establish that the Christian ought to endorse all rights, e.g., the freedom of pornography merchants (cf. Quinn, “Can Christians Be Good Liberals?” pp. 16-18). To do that, Quinn would have to develop a separate argument for each of the liberal rights analogous to the argument he presents that the right to religious freedom is a good that integralist Christians should value.
I want to focus on the branch of Quinn’s argument that suggests religious
integralists should endorse the liberal polity from the perspective of the liberalism of fear.
Quinn presents his argument rather modestly, as simply naming the theoretical
perspective that Wolterstorff proposes as giving support for the liberal polity, hoping that
Wolterstorff will exclaim, paraphrasing the protagonist of Molière’s *Le Bourgeois
Gentilhomme* that he has been a liberal of fear all along. I will let Wolterstorff decide
whether the liberalism of fear adequately represents his position. For now, I want to
explore Quinn’s suggestion that the liberalism of fear might provide a way for integralists
to support liberal regimes.

Wolterstorff thinks that Christians should support liberal policies that respect
people’s dignity and oppose those which denigrate and humiliate people.

Wolterstorff proposes that a violation of a person occurs when
someone carries out a direct and serious attack on or an unwanted
intrusion into “that person’s body, that person’s inner life, that
person’s deep moral and religious convictions, that person’s deep
investment in the world.”

Boiling Wolterstorff’s liberalism down to its simplest terms, we
may say that it places most of the weight in the argument for the
liberal polity on its ability to prevent, or at least to mitigate, the
great evil of violations of personhood. It counts the liberal polity’s
ability to secure the freedom to choose a comprehensive doctrine
of the good and to act upon it once chosen as a small good.

This puts Wolterstorff in the camp of the liberalism of fear “broadly construed,”
says Quinn. “Like Wolterstorff’s liberalism, the liberalism of fear argues for the liberal

\[\text{Cf. Quinn, “Can Good Christians Be Good Liberals?” p. 27.}\]

\[\text{Ibid., p. 19.}\]

\[\text{Ibid., p. 21.}\]
polity, not from some great good it aims to secure for citizens, but from a great evil against which it shows promise of protecting its citizens.”

Christians ought to reject the relativism that sometimes accompanies this strand of liberalism, and will probably reject Montaigne’s skepticism (though they might accept it if they locate him “in the tradition of skeptical fideism that runs from Pascal to Kirkegaard.”) On the other hand, Quinn thinks that, at least in Shklar’s formulation, there is a feature of the liberalism of fear that Christians would have to reject. Shklar thinks that liberals of fear cannot put offences to God ahead of offences against man. “To hate cruelty with utmost intensity is perfectly compatible with Biblical religiosity,” she argues, “but to put it first does place one irrevocably outside the sphere of revealed religion” because it puts the love of neighbor ahead of the love of God. Quinn argues, persuasively, that this either/or rhetoric is unnecessary: if you put offences to man at the head of your secular concerns, that is sufficient for justifying “the core values” of the liberalism of fear.

In particular, I think Christians of Wolterstorff’s stripe, who see God’s image in human persons, can hate cruelty because acts of cruelty are direct and serious attacks on the personhood of human beings. They can hold that cruel acts are both sins, because they offend God, and grave moral wrongs, because they are assaults on images of God. And they can even go on to say that such acts are offensive to God precisely because they are violations of the divine image in persons. …I reckon that ranking cruelty first among the

---

25 Ibid., p. 22.
26 Ibid., p. 24.
27 Ibid., p. 25, citing Shklar, Ordinary Vices, pp. 8-9.
ordinary vices and second only to pride among all the vices would be good enough to underpin a robust commitment to the core values of the liberalism of fear.

Quinn’s purpose is not to develop the liberalism of fear in any detailed political way, but simply to show how the concerns of this strand of liberalism overlap with the Christian’s (and Jew’s and Muslim’s\(^2\)) concerns about defending human dignity. I want to pursue this suggestion by examining the liberalism of fear more directly. For if the liberalism of fear can lead to a justification of the liberal polity that is acceptable even to important categories of integralists, then perhaps that will lay to rest the worries that liberalism will be unstable under conditions where integralism is widespread.

2.4 The Humane Basis of the Liberalism of Fear

The liberalism of fear is the name Judith Shklar gives to a tradition of liberalism grounded “in the conviction of the earliest defenders of toleration, born in horror, that cruelty is an absolute evil, an offence against God or humanity.”\(^3\) The liberalism of fear “makes cruelty the first vice,” offering an alternative to “cruel military repression and violence.” Shklar argues that it is “by definition not to be forced or even promoted by the use of political authority,” suggesting that if it were so forced or promoted it would “undermine its ethical structure.”\(^3\) \(^1\) It is “a strictly political theory…not necessarily linked

\(^{2}\) Cf. ibid., p. 19: “It seems to me that the thought of persons being images of God, which comes from the Hebrew Bible, also provides for Jews and Muslims a reason for regarding violations of persons as great evils.”


\(^{3}\) Idem., Ordinary Vices, p. 5.
to any one religious or scientific doctrine, though it is psychologically more compatible with some rather than with others.”

Although this liberalism will reject some traditions, it “need not decide among traditions that are not hostile to its aspirations, nor does it have to regard the claims of any traditions as inherently false, simply because it does not meet scientific standards of rational proof.” It is intellectually modest, because it does not need to make grander theoretical claims than that cruelty is the *summum malum*. It is “entirely non-utopian,” worrying about specific ways of reducing cruelty in the here and now, rather than developing a vision of the society without cruelty toward which all societies should tend. In this it is different from the “liberalism of natural rights” advocated by Locke and from the “liberalism of personal development” advocated by Mill. These are “of course perfectly respectable expressions of liberal doctrine,” but they lack the historical and empirical approach, the concern with controlling forces of actual oppression, characteristic of the liberalism of fear.

The liberalism of fear is not a contract doctrine, seeing instead the basic units of political life as “the weak and the powerful.” Governments are inevitably stronger than


34 Ibid., pp. 26-7.

35 This claim is mine, and requires some qualification. Shklar does think that the liberalism of fear can be the result of a consensus that cruelty is the worst thing we do to one another, and so her liberalism does share with contractarian theories an emphasis on consensus. But Shklar’s theory does not attempt to argue from an idealized person to an idealized politics, but rather to scrutinize politics as it is to uncover and guard against potential sources of cruelty. Hence her claim, “For this liberalism, the basic units of
their citizens, and so are always potential dangers to them. The concerns of Montesquieu and Madison about separating the powers of government so that they check and balance each other are manifestations of this liberalism. This liberalism is not reducible to Berlin’s negative liberty, which is theoretically distinct from any political institutions. Shklar writes,\textsuperscript{36}

[Berlin’s] very clear demarcation of negative liberty is the best means of avoiding the slippery slope that can lead us to its threatening opposite. Nevertheless, there is much to be said for not separating negative liberty from the conditions that are at least necessary to make it possible at all. Limited government and the control of unequally divided political power constitute the minimal condition without which freedom is unimaginable in any politically organized society. It is not a sufficient condition, but it is a necessary prerequisite. No door is open in a political order in which public and private intimidation prevail, and it requires a complex system of institutions to avoid that. …Socially that also means a dispersion of power among a plurality of politically empowered groups, pluralism in short, as well as the elimination of such forms and degrees of social inequality as expose people to oppressive practices. Moreover, there is not particular reason to accept the moral theory on which Berlin’s negative freedom rests. This is the belief that there are several inherently incompatible moralities among which we must choose, but which cannot be reconciled by reference to a common criterion. …Whatever the truth of this metapolitical assumption may be, liberalism can do without it.

This liberalism does not have one \textit{summum bonum} toward which all moral theories are subordinate, but it does have a \textit{summum malum} away from which all moral theories should flee—a sort of negative overlapping consensus that cruelty should be avoided and

\begin{quotation}

political life are not discursive and reflecting persons, nor friends and enemies, nor patriotic soldier-citizens, nor energetic litigants, but the weak and the powerful,“ (ibid., p. 27).
\end{quotation}

\textsuperscript{36} Ibid., pp. 28-9.
minimized wherever possible. This is the only universal claim of the liberalism of fear, thinks Shklar: “Because the fear of systematic cruelty is so universal, moral claims based on its prohibition have an immediate appeal and can gain recognition without much argument.”

The instinctive aversion to systematic cruelty has to be supplemented by moral claims, though; instincts are particular, and do not lead me to fight for the protection of others where there is no reason for me to worry about being oppressed myself. So the liberalism of fear has to be supplemented by some moral claim that leads us to outlaw cruelty for others as well, based on some moral principle. Shklar insists, though, that so long as there are moral theories to do this—she thinks Kantianism and utilitarianism meet this standard—then liberalism need not endorse any particular supplementary moral theory.

Shklar thinks putting cruelty first makes several requirements of liberalism. Governments must be allowed when necessary to use the threat of punishment and coercion to prevent greater cruelties, even though such coercion is (or can be or risks becoming) cruel. Because it is dangerous to concentrate coercive power in the government, liberalism imposes several duties regarding it. The state and its agents must be treated with suspicion. All their acts must be transparent and subject to public review. There must be division of power to create checks and balances. There must be strong and vibrant voluntary associations, so that civil society balances and affects the use of

---

37 Ibid., p. 30.

38 Ibid. As Rawls observes, this feature of Shklar’s liberalism of fear makes it a kind of political liberalism. Cf. PL p. 374n1.
coercive power. There must be a *distinction between the public and the personal spheres*, and the personal sphere is the area of *toleration*, “the irreducible limit on public agents.” How to draw the line is a matter for the particular supplementary theory, but that there is a private personal sphere immune from government concern is an essential characteristic of liberalism. The private sphere should include some sort of *right to property*, “an indispensable and excellent way of limiting the long arm of government and of dividing social power, as well as of securing the independence of individuals.” This right to property is not natural but legal. It is in the service of the public interest, and so is limited by that end. There must be established and well-understood *legal procedures* that limit the ability of any representative of the state to intimidate citizens arbitrarily. The liberalism of fear must pay close *attention to the social and political conditions* that actually obtain, “in order to act here and now to prevent known and real dangers.” Liberalism must honor the *rule of law*, “the prime instrument to restrain governments.” It also requires that *laws and regulations be limited* by concerns for safety, so that the areas in which one can transgress the law are limited by necessity. Therefore, the liberalism of fear must defend *equal rights and their legal protection*, as well as the *institutions of representative democracy* and of an accessible, fair, and

---

39 Ibid.

40 Ibid., p. 24.

41 Ibid., p. 31.

42 Ibid.

43 Ibid., p. 36.
independent judiciary that is open to appeals.\textsuperscript{44} Finally, the liberalism of fear has universal scope, because once we put the fear of cruelty and of fear at the center of our political theory, we should be inclined “to look with critical attention to the practices of all agents of all governments and to the threats of war here and everywhere,”\textsuperscript{45} and not just be concerned about our own parochial interests.

Shklar’s version of liberalism purports to be less theoretically burdensome than even political liberalism. Rather than trying to defend liberal rights or institutions directly by reference to abstract principles or idealized social contracts, she appeals to the universal instinctual aversion to systematic fear as a way to ground liberal commitments. Her view approaches consequentialist justifications of liberalism when it emphasizes reducing actual cruelty here and now, but it differs in that it does not attempt to define happiness or even to tell the citizens to pursue it rather than, for example, salvation.\textsuperscript{46} It approaches Kantian justifications of liberalism when it places human dignity at the forefront of politics, but it pulls back from giving anything more than the thinnest of accounts of human dignity, roughly “that which is degraded by humiliation and fear.” Shklar thinks this liberalism requires a certain virtuous moral character to be widespread,

\begin{footnotes}

\footnote{Ibid., p. 37. Reflecting on the role of democratic institutions in preventing state cruelty, Shklar quips, “It is therefore fair to say that liberalism is monogamously, faithfully, and permanently married to democracy—but it is a marriage of convenience.”}

\footnote{Ibid.}

\footnote{Ibid. p. 31.}

\end{footnotes}
although she insists, “it is not the task of liberal politics to foster [virtuous characters] simply as models of human perfection.”

Let us recall that Quinn thinks the liberalism of fear can be attractive for Christian integralists because it has so few theoretical commitments, and thus fewer occasions to conflict with Christian doctrines. If Shklar is right, then citizens can support liberal institutions if they can accept cruelty as the *summum malum* of politics, which Quinn plausibly argues is at least an available option for Christians, if not an outright imperative. If we take Shklar’s view to be simply that a liberal is one who puts human dignity first, then it is clear that a good Christian can and probably ought to be a good liberal. But there is reason to think this definition of liberalism is too broad. Most political conservatives operate out of a concern for human dignity, for example. If we grant the racist his doubts about the humanity of lesser races, then we could imagine a certain sort of ‘liberal’ racist who identifies human dignity with the dignity of a few races, and puts the protection of that dignity first. As Jacob Levy puts it, “There is more content to liberalism than *only* fear of cruelty, because the fear of cruelty does not uniquely dictate *liberal* politics. … A different social theory might generate a republicanism of fear, or a conservatism of fear, or a socialism of fear. These might not, …

Ibid., p. 34. Shklar’s *Ordinary Vices* is one attempt to give a sketch of the moral virtues essential to support the liberalism of fear. In “The Liberalism of Fear,” she claims that “there is a very clear account of what a perfect liberal would look like more or less. It is to be found in Kant’s *Doctrine of Virtue*, which gives us a very detailed account of the disposition of a person who respects other people without condescension, arrogance, humility, or fear. He or she does not insult others with lies or cruelty, both of which mar one’s own character no less than they injure one’s victims” (pp. 33-4). She also lists “moral courage, self-reliance, and stubbornness to assert themselves effectively” as liberal virtues (p. 33).
for example, accord the primacy to *state* cruelty and *political* violence that Shklar does.**48**

Rawls argues that for a regime to be stable, it needs to operate from *principles* that are widely accepted. As we saw, Shklar admits that our merely instinctive aversion to fear is not stable enough; therefore, the liberalism of fear requires supplementary moral theories that make the prevention of cruelty a core principle of universal moral laws. This suggests that there are several distinct stages to her liberalism of fear. First there is the decision to put avoiding violations of human dignity at the center of politics. Next, there needs to be a moral theory that elevates this sentiment to a moral principle and gives content to the notion of human dignity by defining the *summum malum*. Then, there needs to be an analysis of the greatest challenges to human dignity in a particular society, which will probably imply some further theoretical work to define by what standard one decides which challenge is greatest. Then, there needs to be a political theory about how to design institutions and practices to defend human dignity against such challenges. Then, there needs to be a further theory about how to reason morally and politically within the parameters laid down by these just institutions.**49** There may even need to be a


**49** See Levy, *The Multiculturalism of Fear*, pp. 34-37 for a slightly different version of this analysis.
theory of something like rights to supplement this liberalism, despite its pretensions to
develop liberalism without rights.  

\[\text{Diagram: The Liberalism of Fear:} \]

\[\text{Possible Stages of Articulation} \]

- Fear of Cruelty
- Utilitarianism
- Habermasian Liberalism
- Kantianism
- Rawlsian Liberalism
- Communitarianism
- Raz’s Perfectionism
- Justice as Fairness (TJ)
- Political Liberalism

The figure above shows how Rawls’s justice as fairness and political liberalism can plausibly be seen as versions of a fully articulated liberalism of fear. As we have seen above, Rawls thinks that a person’s sense of self-worth is a good to be protected above all else, and to be violated only when it is absolutely necessary to protect the self-respect of others. He understands his project to be an improvement over utilitarian theory, which could not in principle rule out the justice of diminishing the welfare of one person

\[\text{50} \text{ Levy points out that Avishai Margalit’s attempt to develop a version of the liberalism of fear built around avoidance of humiliation in } \text{The Decent Society} \text{ (Harvard University Press, 1996) still relies on a theory of rights at some key points (Levy, op. cit., p. 37; cf. Margalit, op. cit. p. 141). Shklar draws out the connection between the liberalism and fear and the demand for political rights in “Injustice, Injury, and Inequality: An Introduction” in } \text{Justice and Equality Here and Now} \text{. Frank S. Lucash, ed. (Cornell University Press, 1986) pp. 13-33, esp. p. 25ff.}\]

\[\text{51} \text{ As noted above, Rawls sees Shklar’s liberalism of fear as a version of political liberalism. See PL 374n1.}\]
if it would maximize the utility of the society as a whole.\(^5\) Rawls thinks the basic liberties provide the best protection for the social bases of self-respect, and that the conception of justice as fairness best guarantees the basic liberties.\(^3\) Within the well-ordered society, the maximin principle should be used to allocate the primary goods, and it is the job of the just institutions of the basic structure to protect the ability of citizens to carry out their comprehensive conceptions of the good so long as doing so does not take away from the self-respect of other citizens. Rawls requires that all fundamental political decisions be public and transparent, and that they are restricted to the public realm. He endorses toleration within the private sphere guaranteed by the rule of law, and a strong, independent judiciary to protect it. In almost all particulars, justice as fairness fulfills Shklar’s list.\(^4\)

The affinities between Rawls’s liberalism and Shklar’s are not accidental—they were close colleagues at Harvard for two decades.\(^5\) As Quinn’s account suggests, at its earliest stages of articulation, Shklar’s negative liberalism can be combined with a wider variety of comprehensive views than can Rawls’s political liberalism. At those earliest

\[^{52}\text{Cf. TJ §§29-30.}\]

\[^{53}\text{Cf. PL 319: “Since only the two principles of justice guarantee the basic liberties, they are more effective than the other alternatives in encouraging and supporting the self-respect of citizens as equal persons.”}\]

\[^{54}\text{The one important difference is that Shklar’s view is concerned with preventing actual instances of cruelty “here and now,” while Rawls is concerned with “ideal theory,” at least explicitly.}\]

\[^{55}\text{Indeed, Rawls acknowledges Shklar’s collaboration with him in the Introduction to PL xxxiv. Shklar likewise acknowledged her intellectual debt to Rawls in “Injustice, Injury, and Inequality,” p. 13.}\]
stages, however, Shklar’s liberalism is more likely to be “political in the wrong way,” i.e., based on compromises with on-the-ground facts about power relations which may or may not be just. As we have seen, Rawls thinks that the problem of stability requires that a conception of justice be based on moral principles which all citizens internalize; otherwise, it is a mere modus vivendi, which might disintegrate if the political circumstances change significantly. And as we have also seen, Shklar agrees that some sort of moral principles are necessary to ensure that opposition to cruelty at all times everywhere is central to politics. Rawls’s theory is not the only one that provides moral principles for opposing cruelty, but it does give us a sense of what a full-fledged liberalism of fear could look like.

But what about Quinn’s suggestion that the liberalism of fear might provide even integralists with reasons for endorsing liberalism? An integralist, who wants to organize even the political community around his comprehensive doctrine, might be able to agree with Shklar’s basic claim that cruelty ought to be avoided. But as the liberalism of fear develops through its later stages of articulation, there might come a point beyond which the integralist cannot go. He cannot go all the way to Rawlsian political liberalism, as we have seen. If Shklar is right that avoiding cruelty requires that there be a distinction between the public and private spheres and that a policy of tolerance apply to the private

56 Cf. PL 40.


sphere, then the integralist will have to be especially careful about how this sphere gets defined. If there is a way to define these spheres to permit the integration of one’s life around a comprehensive doctrine, then the integralist can be a liberal. But if not, then the integralist is presented with a stark choice: either reject integralism, or reject the premise that cruelty is the *summum malum*. Needless to say, for every citizen who is prepared to use cruelty to advance his or her comprehensive doctrine in the political realm, or is at least open to this possibility, the stability of the liberal regime decreases. And if there are likely to be many such people, then liberalism itself is unstable.

So what we called the imperative of stability should drive us to find moral principles that even the integralists can accept. Such principles should not be political in the wrong way, as they would be were they a mere accommodation to the fact that integralists exist, or to the fact of their political power. They must be able to serve as grounds for a stable overlapping consensus for a perpetual society, and so cannot depend on the particular balance of power of one era. These principles, however, must be able to accommodate both the integralists and the liberals, a feat which, if possible, would require a very careful parsing of the public/private distinction. Let us look at this distinction more closely, to find out why it is deemed necessary, and what are its implications.
CHAPTER THREE

THE PUBLIC, THE PRIVATE, & THE BURDENS OF JUDGMENT

3.1 The Public and the Private

Avishai Margalit argues in *The Decent Society* that some form of privacy is a basic human need. It may manifest itself in different forms (Eskimos who spend the whole winter together in an igloo express it in one way, Howard Hughes in another), but there is always some “minimal sphere for individuals’ control over their own interests.”  

This might account for our tendency to say that we *respect* the privacy of others, as if their need for privacy is a fact existing independently of our wishes. Margalit even suggests that violating another’s privacy is a paradigmatic act of humiliation, the *summum malum* in his version of the liberalism of fear, and that constant invasions of privacy are characteristic of totalitarian regimes.

If privacy is such a basic human need, then integralists too will want some sphere of privacy. Their integration of their lives around their comprehensive doctrines will thus have to make some space for a distinction between the personal and the public—they cannot expect their lives to be so totally ‘integrated’ that the personal is political all the

---

1 Avishai Margalit, *The Decent Society*, ch. 12. The quote is from p. 204.
way down. 2 (Or if they do expect this, then we can regard them as pathological. In any case, if privacy is so fundamental a need, the party of those who deny it would not achieve numbers sufficient to cause political instability.) If we recall Quinn’s characterization, integralists refuse to differentiate their lives in one or more of the ways (e.g., into separate spheres for economy, politics, religion) that modern societies tend to require. Their trouble with many versions of liberalism is not the mere fact that they require some sort of distinction between the public and private spheres, but that they demand a separation between citizens’ public activities and their comprehensive doctrines in some area (varying with the flavor of integralism) that the integralists find illegitimate and unacceptable.

If we want to bring the integralists to accept liberalism, we ought to reconsider whether liberalism can do without the features the integralists find most objectionable. As we argued in the previous section, how liberalism divides the public and private spheres is a key issue: an integralist who is prohibited from incorporating his comprehensive doctrine into his public actions is likely to become alienated from the liberal regime and to withdraw his support for it. So in this section, we will examine some canonical arguments in favor of excluding comprehensive doctrines from political life, to see whether these can be altered to garner the support of integralists, or whether they have to

2 In a helpful treatment of this topic, Norman Daniels points out that his rabbi will often avoid partisan political topics in sermons, because reasonable people can disagree about politics (an ironic parallel to the motivation for public reason!). “[Rawls] should not, then, be criticized merely for drawing a boundary that bifurcates justificatory practices. We do that commonly and for good reason at many points in our lives, both in private and public contexts. The challenge [to the integralist] must be more specific if it is to be serious…” See “Reflective Equilibrium and Justice as Political” in Davion and Wolfe, pp. 148-50.
be rejected altogether. We will look at the *locus classicus* for this distinction in John Locke, then at the arguments Rawls gives for the distinction, and finally at how Shklar draws the line. After sifting the strengths and weaknesses of these views, always keeping in mind our twin goals of stability and humanity, we will try to determine whether liberal principles can create a stable regime for a pluralist society that might include substantial numbers of integralists.

3.2 Locke: The Care of Souls is a Private Affair

Locke’s first *Letter on Toleration* gives the classic arguments for the distinction between the public and the private spheres:

The care of souls is not committed to the civil magistrate, any more than to other men. It is not committed unto him, I say, by God; because it appears not that God has ever given any such authority to one man over another as to compel anyone to his religion…. In the second place, the care of souls cannot belong to the civil magistrate, because his power consists only in outward force; but true and saving religion consists in the inward persuasion of the mind, without which nothing can be acceptable to God. And such is the nature of the understanding, that it cannot be compelled to the belief of anything by outward force…. These considerations, to omit many others that might have been urged to the same purpose, seem unto me sufficient to conclude that all the power of civil government relates only to men’s civil interests, is confined to the care of the things of this world, and hath nothing to do with the world to come.

---

3 John Locke, “A Letter Concerning Toleration,” in *The Works of John Locke*, (1823; reprint, Scienta Verlag Aalen, 1963) v. 6, pp. 10-13. I omit Locke’s third argument that the care of souls is not committed to the civil magistrate, an argument about the absurdity of the doctrine of *cuius regio, eius religio*, since is not relevant to our immediate purposes.
Locke’s argument for religious toleration is different from his contractarian argument in his *Second Treatise* for the rights to life, liberty, and property, but the conclusions are similar. There Locke writes, ⁴

> But though men when they enter into society give up the equality, liberty, and executive power they had in the state of Nature into the hands of the society, to be so far disposed of by the legislative as the good of the society shall require, yet it being only with an intention in every one the better to preserve himself, his liberty and property (for no rational creature can be supposed to change his condition with an intention to be worse), the power of the society or legislative constituted by them can never be supposed to extend farther than the common good, but is obliged to secure every one's property by providing against those three defects above mentioned that made the state of Nature so unsafe and uneasy. And so, whoever has the legislative or supreme power of any commonwealth, is bound to govern by established standing laws, promulgated and known to the people, and not by extemporary decrees, by indifferent and upright judges, who are to decide controversies by those laws; and to employ the force of the community at home only in the execution of such laws, or abroad to prevent or redress foreign injuries and secure the community from inroads and invasion. *And all this to be directed to no other end but the peace, safety, and public good of the people.*

Because the people who were born free give up their rights for a purpose, the power of the state cannot legitimately extend beyond the state’s original charter. Those powers concern in the words of the First Letter, “only men’s civil interests…the care of the things of this world.”⁵ Locke argues that citizens do not give the state absolute power to govern them, but only grant it those powers necessary to secure the common goods of

---


⁵ Although we cannot go into them, there are some difficulties in deriving a consistent doctrine from these texts; see Jeremy Waldron, *God, Locke, and Equality: Christian Foundations of Locke’s Political Thought.* (Cambridge University Press, 2002) pp. 208-214.
this world—peace, safety, and the common good of society—and to preserve their own lives, liberty (especially religious liberty), and property. Locke’s argument grounds the distinction between public and private spheres on a division of labor—the state has been assigned certain duties and has been given sufficient power to carry them out, so it ought not go beyond those responsibilities.

Locke’s argument for excluding the state from attempting to govern the inner life of its citizens would not be acceptable to integralists. For example, Jeremy Waldron points out, Locke claims that people could not consent to a theocracy or an absolute monarchy, and that they cannot alienate their rights, but he isn’t clear about why.\(^6\) Integralists who would not want to accept Locke’s description of the public and private realm would be under no rational obligation to do so.

Interestingly, Rawls and most other liberals would not accept that the state has no interest in the souls of its citizens, at least not to the absolute degree that Locke advocates in the first *Letter*. As Joshua Cohen notes, Rawls is able to argue from liberal political procedures to substantively egalitarian positions because he has a morally substantive view of persons, “a conception of the properties of human beings that are important for the purposes of political justification.”\(^7\) Of course, any social contract theory has some

\(^6\) Waldron, *God, Locke, and Equality*, p. 214: “What would be the objection to a group of people banding themselves together in an all-purpose association, since they were entitled to band together voluntarily in a state and entitled also to band together voluntarily in a religious association? Clearly something had to give…”

view of persons, but Rawls has a view of persons that extends to moral psychology, the area of human life that Locke placed off limits from the civil magistrate.

Rawls argues both in TJ and in PL that for a regime to be stable over time the citizens must acquire certain moral characteristics so that they will tend to endorse the principles of the regime and work to uphold them. While almost all citizens have at least some degree of the requisite moral powers at birth (perhaps only the mentally handicapped or mentally ill do not), some people—the unreasonable—fail to exercise them. If enough people fail to exercise their moral powers, then the regime will not be stable. For these reasons, Rawls and many contemporary liberals make some provision for civic education in the principles of liberalism.  

3.3 Rawls on the Original Position and Public Reason

We have already established that Rawlsian political liberalism is prepared to treat certain ‘unreasonable’ people harshly, and that the integralists make up a substantial portion of the unreasonable because they refuse to separate their comprehensive doctrines from their public political activity. Now we need to discover why Rawls thinks

---


9 There might be some people who are unreasonable without being integralists—a criminal or would-be tyrant might refuse to exercise her sense of justice, but out of a lust for the power to dominate others, and not out of a principled decision to organize her life around a comprehensive doctrine. My argument in this section does not attempt to extend liberalism to all the unreasonable, but only to integralists.
such features are necessary. We saw (and applauded) that Rawls built justice as fairness to protect the social bases of self-respect. Rawls goes on to specify that self-respect is dependent on a certain view of autonomy, where it is a violation of autonomy (and thus self-respect) for the state to coerce a person without being able to explain the grounds for the coercion in terms that the person will regard as reasonable. Rawls does not argue for this view, other than to claim that it is a value implicit in the culture of liberal democratic societies.\(^{10}\) He builds this view of autonomy into his original position, from which he derives his idea of public reason.

The original position is a device meant to help us structure our beliefs about politics so that the categorical imperatives of morality are more fundamental than and prior to any hypothetical imperatives regarding the pursuit of private conceptions of the good.\(^{11}\) The features of the original position are supposed to model what we actually believe to be the conditions of morality and of moral reasoning.\(^{12}\) If someone rejects one of Rawls’s major conclusions, he should be able to point out some feature of the original position that he is unwilling to accept, or else find some error in Rawls’s deductive argument proceeding from the original position.

\(^{10}\) Cf. Paul Weithman, *Religion and the Obligations of Citizenship* (Cambridge University Press, 2002), ch 7. Weithman concludes that a certain positive view of autonomy and self-worth undergirds Rawls’s characterization of public reason, and that it can be reasonable to reject that view of autonomy. See p. 222.


\(^{12}\) Cf. JAF 80 where Rawls twice emphasizes that the original position models what we “here and now” believe.
There may be many points at which the integralist will object to Rawls’s argument, but the most fundamental objections (and the most difficult to rebut) will be objections to features of the original position. The original position is supposed to model the fair conditions of agreement among citizens as free and equal (hence, the two principles), and it is supposed to model the acceptable restrictions on reasoning within those fair conditions (hence, the principles of public reason). An integralist can object to both models.

1. He can object to the idea of free and equal persons as prejudiced against integralism. An integralist wants to integrate his life around some set of values or desires. Those values we can call his higher-order interests. It is reasonable to assume that the integralist’s higher-order interests are not those expressed by Rawls’s two principles. For the integralist, then, the interest in autonomy upon which the two principles depend is of a lower order than the interests around which he organizes his life. Organizing his life around his set of values expresses his belief that the values are true and stably true; a ‘right’ to abandon those values and the life organized around them is tantamount to a ‘right’ to abandon the truth—and just as illegitimate. Thus, these higher-order interests will be unconditional, where an interest is unconditional just in case it is a desire that the

---

13 JAF pp. 16-17.

14 Cf. PL, p. 74: “To say that these interests are ‘higher-order’ interests means that, as the fundamental idea of the person is specified, these interests are viewed as basic and hence as normally regulative and effective.”

15 Presumably, any Rawlsian integralists would not be particularly burdened by a liberal regime. See Tomasi, Liberalism Beyond Justice, pp. 17-20.
state of affairs that is its object should obtain whether or not one continues to desire that it obtain. Let us say that an unconditional commitment is a commitment to an unconditional desire. It is clear that the integralist’s commitment to his higher-order values is unconditional in this sense.

Rawls assumes that the desire for autonomy is a higher-order desire for all reasonable citizens, and so claims that the parties in the original position will agree upon the principle of liberty of conscience. His construction of the original position rules out even the possibility that those represented by the parties might have desires for unconditional commitments to their comprehensive doctrine. Stephen White argues, convincingly, that Rawls has to do this for the original position to yield any determinate principles at all. If it is psychologically possible that they might desire to bind themselves once and for all to a set of desires they possess at a certain time—like Odysseus binding himself to the mast so that he could not respond to the sirens—then the parties could not conclude with Rawls that the liberty to revise these desires, what Rawls calls the liberty of conscience, must be a principle of justice. The same problem holds for almost any unconditional desire: if the parties could imagine that they could have an unconditional desire to be an atheist who would never want to be a Christian, or a Christian who would never want to be an atheist or a Muslim, or that they would want to marry without the possibility of divorce, then they could not all agree on the first principle of justice as fairness. For if they could imagine this possibility from behind the veil of ignorance, then

16 This definition is from Stephen White, “Rawls and Ideal Reflective Equilibria” in The Unity of the Self (MIT Press, 1991) p. 303. The argument here closely follows this excellent article.
they would recognize that they would be in frequent opposition to a society in which the principle of liberty was fundamental. In such a society, they would see, the strains of commitment would be too much to bear.\textsuperscript{17}

As White notes, the only way Rawls can save his argument is by stipulating that the parties in the original position have a highest order unconditional commitment to being able to revise their comprehensive doctrines. (This is one way of characterizing his stipulation that political liberalism depends on a majority of its citizens being reasonable and embracing a sense of justice, i.e., liberal democratic citizens who value a certain understanding of autonomy.\textsuperscript{18}) But this is tantamount to abandoning his claim that our views about justice can be specified independently of our comprehensive doctrines, a claim White refers to as the \textit{justificatory sense} in which the right is prior to the good.\textsuperscript{19} As White writes:\textsuperscript{20}

\begin{quote}
If a subject in the original position must know that the desire for revisability (or more directly, a desire for a society governed by the
\end{quote}

\textsuperscript{17} One might argue that White’s argument requires that the parties have not only unconditional desires, but unconditional desires with consequences for the basic structure. It is conceivable that someone might have an unconditional desire to be a Muslim herself, but not an unconditional desire that all citizens live under Islamic law. I suspect that all integralist views will have a public dimension, but if not, then the argument should be amended in this way. I thank Paul Weithman for this point.


\textsuperscript{19} See White, \textit{The Unity of the Self}, pp. 287-8: “Let us say that the right is prior to the good in the justificatory sense if our commitments where rights are concerned can be justified independently of our substantive commitments concerning the good. Communitarians…disagree with Rawls’s contention that [the right] has priority in the justificatory sense.”

\textsuperscript{20} Ibid., p. 316.
conception of rights reflected in the first principle) is a part of his or her [highest-order unconditional commitments], then it has been implicitly conceded that in the justificatory sense the good is prior to the right.

Sandel makes a similar point to White, though less rigorously, in the second edition of *Liberalism and the Limits of Justice*. Both grant Rawls the idea of reflective equilibrium, but not the original position or the priority of the right over the good. (Indeed, using Rawlsian terms, one might characterize the current stage of the liberal-communitarian debate as a debate over whether the original position adds anything to the idea of reflective equilibrium, or is simply a method for articulating some views that certain liberals want to balance against their other views in reflective equilibrium.) White’s argument is the stronger of the two, since it shows that the very possibility of unconditional commitments makes it impossible for the parties of the original position to arrive at consensus on Rawls’s principles of justice, or any other principles that depend on autonomy or the ability to revise one’s comprehensive doctrine. Rawls’s decision to limit his theory to reasonable citizens allows him to evade this objection, but as we have argued in earlier sections, this comes at unacceptable costs: it treats some citizens inhumanely, and it leads to political instability.

---


22 White (*The Unity of the Self*, pp. 320-1) is emphatic on this point: “The problem is not that the original position yields a pair of determinate principles in virtue of an argument that begs the question against a large class of nonliberals. If I am right, on the assumption that unconditional desires are not irrational, the original position yields no determinate principles….The upshot is not that we cannot be rationally motivated to respect the rights of others, that a society must be ruled by force, or that it must consist in members who share exactly the same conception of the good. The conclusion is that our commitments to the rights of others must have their support in our substantive desires, and thus that the right is not, in the justificatory sense, prior to the good.”
2. As we have seen, an integralist need not accept the view of autonomy that undergirds Rawls’s first principle of justice. The integralist can also reject Rawls’s idea of public reason. In part, this second rejection is a consequence of the first, since public reason is a principle that depends on the political conception of the citizen as autonomous. What Rawls calls ‘the duty of civility’ requires that citizens be confident that their political actions regarding fundamental political questions can be shown to be derived from basic political values. This duty of civility derives from the idea that free and equal citizens retain their autonomy only if whatever coercive action taken against them is based on reasons that they can accept. Therefore, coercion, to be legitimate under the conditions of reasonable pluralism, must be based on reasons that are publicly evident and to which everyone can agree is reasonable. If such reasons can be found, then the conditions of social and political cooperation between the free and equal citizens will be stable. Rawls calls this ‘public justification,’ or ‘publicity.’ He thinks something like this requirement

---

23 PL p. 217. See also “The Idea of Public Reason Revisited,” in *Law of Peoples*, p. 144 for what Rawls calls ‘the proviso’, which allows one to introduce one’s comprehensive doctrine into political discussion provided that it be possible “in due course” to give properly public reasons for the same policies and principles. (NB: I am amending Rawls’s view along the lines suggested by James Sterba, “Rawls and Religion,” in Davion and Wolf, pp. 42-43. Rawls says that each citizen has an obligation to meet the duty of civility and the proviso when acting on fundamental political matters. Sterba suggests that the duty of public justification is a collective duty of those advocating a position, rather than an individual one, and that this duty is discharged if one person can make a case for the position in terms acceptable to public reason. “From time to time,” suggests Sterba, “some well-placed majority advocates help make accessible sufficient…reasons for the minority to go along with the will of the majority.” I think the spirit of the proviso, that legitimate coercion must be reasonable in the eyes of the coerced, is satisfied by Sterba’s friendly amendment.)

24 Or at least agree is not unreasonable (cf. PL p. 253). See also PL p. 390; p. 136.

25 For publicity, see PL §II.4, I.6.2; TJ §29.
is intrinsic to any contract theory, and contrasts it with the heteronomous claims of aristocracy or autocracy.

On this view, it follows that if some basic political decisions are not likely to be intelligible and acceptable to all citizens, then they do not meet the publicity condition, and are therefore illegitimate. Thus, the legitimacy of the liberal regime depends on there being some principles or policies that it is possible for all citizens to find intelligible and acceptable. If such principles are not to be found, then either the standard must be abandoned or no regime can be legitimate. These considerations account for why Rawls frames the fundamental question of Political Liberalism in terms of a search for the conditions of the possibility of a legitimate regime given the fact of reasonable pluralism: “How is it possible for there to exist over time a just and stable society of free and equal citizens, who remain profoundly divided by reasonable religious, philosophical, and moral doctrines?”

As we saw just above, Rawls achieves a consensus on the two principles of justice by placing artificial constraints on the outcome of the process of reflective equilibrium. Those constraints require him to exclude a priori all those who will not find the claims of liberalism to be intelligible or attractive. Removing those constraints makes it possible

---

26 TJ §3, p. 15, §29.

27 See PL 212-13. Cf. Waldron, “Theoretical Foundations,” pp. 128, 146-50, for the argument that this ‘transparency’ requirement is fundamental to liberalism: “Society should be a transparent order, in the sense that its workings and principles should be well-known and available for public apprehension and scrutiny” (p. 146). “Liberals demand that the social order should in principle be capable of explaining itself at the tribunal of each person’s understanding” (p. 149).

28 PL, p. 4. Emphasis added.
that citizens could have fundamental disagreement about basic political issues, as we have seen. But it also makes it possible that they might find the arguments of their fellow citizens regarding fundamental issues to be opaque or inaccessible to them, independently of whether they agree with them or not. The original position was supposed to ensure that the parties would meet the publicity requirement, so that all political discussion could take place with common premises and a common principle of public reason.\(^{29}\) Its failure to yield determinate principles makes it possible—even likely, in conditions of pluralism—that there will not be a common set of premises or an agreed upon method of public reason.

The presence of the integralists forces us to consider two sorts of disagreement that would not be possible if liberalism could exclude the unreasonable. If all citizens agreed on principles of political justice and public reason, then, Rawls argues, the exercise of practical reason in conditions of freedom would not lead us to accept diverse principles of justice. Within his ideal theory, any disagreement is reasonable disagreement, i.e., disagreement among reasonable persons, who can agree on the principles of political liberalism.\(^{30}\) When these people disagree, it is in areas other than the fundamentals of political justice. Such disagreements are the result of what Rawls calls ‘the burdens of judgment’—the fact that there are all sorts of applications of our theoretical and practical reason where there is no clear answer about how to assess

\(^{29}\) For the connection between the original position and transparency/publicity, see Paul Weithman, “Waldron on Political Legitimacy and the Social Minimum” *Philosophical Quarterly* 45 (1995) pp. 218-224, especially 219-221.

\(^{30}\) Cf. PL, p. 55.
evidence, or balance principles, or handle tough cases, or apply universal norms in particular cases. In these cases, citizens are forced back on their own experiences and judgments.  

Although Rawls doesn’t say this, one might say that the cases where there is no obviously right approach to reasoning are those in which one must apply the method of reflective equilibrium to arrive at a solution.

Now, when one removes the requirement that all citizens are reasonable, there is a likelihood that considerations of basic justice will also be subject to the burdens of judgment, so that people can disagree about how to weigh the different values that Rawls thinks are fundamental to liberal democracies. An experienced politician might weigh the political values differently than a college radical, for example, and each might not be able to explain his decision to the other satisfactorily. John Henry Newman tells of an experienced judge who gives a young colleague the advice that he “always lay down the law boldly, but never give his reasons, for his decision was sure to be right, but his reasons sure to be unsatisfactory.” This is the first sort of disagreement that the presence of the integralist makes possible.

It also becomes possible that citizens might possess incommensurable comprehensive doctrines about politics. This is not the same as simply arriving at different political judgments. Rather, it entails having different internally coherent reasons for a judgment. The political turn in Rawls’s thought assumes that

See PL §II.3.


incommensurable comprehensive doctrines of the good can be reconciled so long as they are all reasonable comprehensive doctrines. But if we remove the assumption that all citizens are reasonable in Rawls’s sense, then we are no longer able to assume that the doctrines will overlap at all.  

The integralists can thus reject the idea of autonomy and the idea of public reason that are at the basis of Rawls’s political liberalism. Indeed, the possibility that there might be significant populations of integralist citizens suggests that it would not be stable or humane for a political regime to divide the public and private spheres according to these principles. If that is true, then nobody ought support a liberalism that relies on these principles, unless it can be stipulated that the citizens of the liberal regime always be homogeneously liberal. This not only violates Rawls’s own concerns about the ‘fact of oppression,’ it seems shortsighted and unstable. (Even in Scandinavian countries, perhaps the best candidates for perpetually homogeneous liberal populations, there is always the possibility of an integralist religious revival or an integralist socialist movement.)

These conclusions ought to be somewhat unsettling. For the features we have been criticizing are exactly those that Rawls hoped could allow a regime to be built around a stable consensus rather than an unstable, balance-of-powers modus vivendi.

Stability is the greatest when all citizens can understand and endorse the principles of the

34 In two of his papers ("Kantian Constructivism in Moral Theory", "Social Unity and Primary Goods"), Rawls appeals to an article by Allen Buchanan, "Revisability and Rational Choice," Canadian Journal of Philosophy 5: 395-408 (1975) which argues that such incommensurability is irrational and thus impossible in the original position. White (The Unity of the Self) effectively criticizes Buchanan’s argument, demonstrating that incommensurability is possible. Rawls makes no reference to Buchanan’s article in PL or later.
regime. Only then will the regime be stable through demographic shifts and across
generations. If the regime has no principles, or if its principles are illegitimate when
viewed from certain influential comprehensive doctrines, or if they depend upon a set of
judgments that are controversial or that require a certain expertise and experience, then it
would seem that stability may not be possible. Rawls saw political liberalism as
answering how a stable liberal democracy could be possible in circumstances of
religious, philosophical, and moral pluralism. If it turns out that his proposed solution is
likely to be unstable, the question still remains with us.

It might be possible to salvage some parts of the Rawlsian system that don’t
depend on the problematic aspects of autonomy and publicity. The criticisms of the
original position, for example, leave open the possibility that the parties could agree on
principles that would allow citizens to have both conditional and unconditional
commitments when the veil of ignorance is lifted. Whatever principles these might be, it
is clear that they would not include a higher order commitment to liberty of conscience.
But if some such principles could be found, then the arguments against the utilitarians
and communitarians could be revived. There would still be some moral imperatives that
place moral and political restraints on deliberations in reflective equilibrium. The right
would still be prior to the good, in that people would still be required to agree on some
political principles that would structure and shape their comprehensive doctrines. People
could still be presented with a standpoint from which to view the common good, even
though the content of ‘purity of heart’ would be different than Rawls originally envisioned it.\textsuperscript{35}

It is not obvious, though, that there are such principles. How can there be principles that everyone will accept if political principles are subject to the burdens of judgment? Doesn’t the specter of incommensurable doctrines make any political consensus impossible? If the principles of the regime need not be transparent, if they are not accessible to all citizens, then how can the citizens agree to them? Without consensus, won’t there be groups of citizens that reject the very principles of the regime, just as the integralists rejected political liberalism? We’ll put these questions on hold until later in the chapter.

3.4 Publicity and Privacy in the Liberalism of Fear

Shklar’s argument for a private sphere is a negative argument, in keeping with her focus on the dangers that states pose to human dignity. Limiting government is a focus of her theory, and that entails that the public sphere should be no larger than is necessary to establish and protect the rule of law and the other institutions protecting human dignity. She is not “in favor of weak governments that cannot frame or carry out public policies and decisions,”\textsuperscript{36} preferring strong governments so long as their strength is checked to avoid tyranny. Nor is she in favor of anarchy. Because she wants a government that is

\textsuperscript{35} Cf. TJ, p. 514.

strong but restrained, the liberalism of fear rejects “those political doctrines that do not recognize any difference between the spheres of the personal and the public.”

Because of the primacy of toleration as the irreducible limit on public agents, liberals must always draw such a line. This is not historically a permanent or unalterable boundary, but it does require that every public policy be considered with this separation in mind and be consciously defended as meeting its most severe current standard. The important point for liberalism is not so much where the line is drawn, as that it be drawn, and that it must under no circumstances be ignored or forgotten. The limits of coercion begin, though they do not end, with a prohibition upon invading the private realm, which originally was a matter of religious faith, but which has changed and will go on changing as objects of belief and the sense of privacy alter in response to the technological and military character of governments and the productive relationships that prevail. It is a shifting line, but not an erasable one, and it leaves liberals free to espouse a very large range of philosophical and religious beliefs.

Shklar supports subsidiarity, civil society, freedom of association, and voluntary associations, not out of a Tocquevillian regard for personal enrichment, but primarily as providing “significant sources of social power” that can act as checks on governmental power. She insists that public entities enact their policies and decisions in accord with publicity, deliberation, and fair procedures, because the possibility of secret and hidden acts makes cruelty easier. For these reasons she is also opposed to traditional communities, because the “unspoken and sanctified practices that prevail” are not publicly accountable, and so allow all sorts of cruel and oppressive practices for which the authorities are not accountable. In answer to communitarian and romantic charges

37 Ibid., pp. 24-5.
38 Ibid., p. 30.
39 She cites Michael Sandel.
that this endorses an atomistic view of the self, she responds that the liberalism of fear protects the freedom of association at the private level, but is suspicious of such critics’ “ideologies of solidarity” at the political level. She fears that these ‘ideologies’ redraw the line between the public and private in ways that would have tremendous consequences; it is not at all certain that the protections of liberal society would survive this redrawing.

This response is revealing. Shklar is willing to accept that the liberalism of fear results in an atomistic view of political man. She doesn’t start with this view, but she’s not afraid of atomism if it is needed to avoid cruelty. This is a defensible bullet to bite, perhaps, but some of Shklar’s arguments tend towards this strictest atomism unnecessarily. For instance, it seems fair to say that trust is a necessary part of any social cooperation. If that is true, a politics that emphasizes suspicion, which seeks to foster and perpetuate a general suspicion of others, seems likely to dissolve the bonds of social cooperation, and unnecessarily. I think that we should follow Quinn’s lead here and try to distinguish Shklar’s particular views from the features of the family of theories she outlines, in order to defend the latter.

It seems, for example, that an integralist might have reasons to accept many of the institutions of liberal democracies. We claimed earlier that integralists must defend some

---

40 E.g., Nancy Rosenblum, Another Liberalism (Harvard University Press, 1987).


42 Rawls is quite clear on this. See PL I.3, pp. 15-22.

43 Indeed, it is a deliberate policy of totalitarian regimes to provoke widespread suspicion and mistrust of one’s fellow citizens (though not of the state).
sort of private sphere. It would seem, therefore, that they ought to agree that the private
sphere should be protected from unnecessary government intrusion (with the force of
‘unnecessary’ varying with the scope of the private sphere a particular integralists wishes
to defend). Since, as Shklar argues, many of the institutions of modern liberal
democracies act as checks on such intrusion, an integralist might be able to endorse many
or most of these institutions.  

The integralist might even be able to accept Shklar’s requirement that all public
actions be publicly known and justifiable. Rawls grounded his publicity requirement, as
we saw, on his view of persons as autonomous; the government could not legitimately act
in ways that the people themselves could not understand, because such action would
violate their autonomy and sense of self-worth. Shklar has a quite different reason for her
mandate that all “public policies and decisions [be] made in conformity to requirements
of publicity, deliberation, and fair procedures.” She thinks that “what is to be feared is
every extralegal, secret, and unauthorized act by public agents or their deputies,” and her
publicity requirement has this purpose as its end.

This version of the publicity requirement is more modest than Rawls’s, and is not
vulnerable to the same criticisms. It is not necessary that every citizen be able to
understand the arguments made by public officials for openness to serve as a check on

44 The different integralists might support or oppose different institutions on these
grounds. Someone who wants to organize her politics around the principle that marriage
is indissoluble might have little objection to the requirement that juries in capital cases
must reach a unanimous decision, while an integralist crime fighter might insist on a
lower standard. On the other hand, the crime fighter might not have an opinion about the
ERA. Both could agree, however, that most of the institutions of liberalism are
compatible with his or her brand of integralism.
state-sponsored cruelty. As long as there is a free and competitive press, there will be an incentive for some reporters to develop sufficient expertise in an area to explain the actions of the public officials to the public and make them available for oversight and criticism. There might be some parts of the government, such as the U.S. Federal Reserve Board or certain undercover police and intelligence operations, which will not be open to the same degree as others. The liberalism of fear would be committed to enacting some sort of democratic oversight of these operations, so that they will not be above scrutiny or accountability.

To put the difference crudely, Rawls treats heteronomy as the *summum malum* of his political liberalism, while other liberalisms of fear need not. It is far from self-evident that all forms of heteronomy are cruel or violate our sense of self-worth. Someone might rationally hold that there could be benign forms of heteronomy (if ‘potentially being subject to coercive state action that I cannot explain or endorse from within my comprehensive doctrine’ is our definition of heteronomy). Indeed, Shklar’s sense that the state is something to fear almost assumes that the state will be heteronomous in some sense, in which case the job of liberal politics is to ensure that the heteronomy will be benign.

——

45 William Galston, *Liberal Pluralism* (Cambridge University Press, 2002) cites the Federal Reserve Board as an instance of a government institution that requires autonomy from direct democratic oversight: “There are some public purposes whose effective pursuit requires specialized knowledge and competence that are not widely shared. When the exercise of such expertise is likely to go against the grain of democratic decision making, there may be a case for insulating the experts against the vagaries of democracy” (p. 85).
There is one aspect of Shklar’s commitment to publicity that the integralist might not share. Against the arguments of some multiculturalists, she argues that liberals should be critical of the tribal societies that exist in some parts of the world. Unless the “unspoken and sanctified practices” of these societies are “openly analyzed and appraised” then “there can be no responsible choices and no way of controlling the authorities that claim to be the voice of the people and its spirit.” This seems unnecessarily strong, for there might be some traditions that not everyone can understand, but which do not perpetuate cruelty. The same arguments used against Rawls’s publicity condition could be mustered against this argument—transparency is not a good in itself that outweighs all other values. Shklar, however, makes an independent argument that leads to a much less objectionable conclusion: “The most reliable test for what cruelties are to be endured at any place and any time is to ask the likeliest victims, the least powerful persons…until that is done, there is no reason not to assume that the liberalism of fear has much to offer” the citizens of tribal societies. This test seems much better, since it holds that one must uncover specific cruel practices in traditional societies in order to attack their traditions. One need not be a cultural relativist to note that there are still some reasons for caution in subjecting the traditions of other societies to outside criticism, but Shklar’s claim that torture is always an evil seems hard to dispute. In any case, this second test is not necessarily objectionable to the integralist.


47 Ibid., p. 35. Of course, there have to be assurances that the answers are honest and accurate.
This is a significant conclusion for contemporary policy debates about whether it is appropriate for the state to enlist faith-based-organizations (FBOs) in certain tasks (e.g., drug-rehabilitation, juvenile crime prevention) where they show promise of being more effective than secular agencies.\textsuperscript{48} It seems that Shklar’s worries about traditional societies would also apply to certain FBOs—they treat men and women differently, for example, or they engage in practices that are not transparent to those who do not understand their faith. If this lack of transparency alone is sufficient to treat those organizations with suspicion, then it is clear that the state ought not to employ them or provide them with support. (Indeed, Shklar’s initial argument would seem to suggest that any FBO with non-transparent practices should be subject to state intervention or regulation.) But if the test for state funding is only that an FBO not be cruel to people, then it seems that a liberalism of fear would not need to oppose in principle state cooperation with FBOs.

This leads to a more general point. If we drop the strong requirement that all government actions be based on reasons that are transparent or accessible to all citizens, in favor of the weaker requirement that government actions not be cruel or humiliating, this opens up significant areas in which the state can cooperate with religious and other traditional institutions. Under the transparency requirement, the preferred mechanism for

\textsuperscript{48} The Charitable Choice provisions of the 1996 Welfare Reform Act, sponsored by then Senator John Ashcroft and signed into law by Bill Clinton allow for such cooperation between federal agencies and FBOs. There is an extensive and growing literature on this topic. See especially the collection Derek Davis and Barry Hankins, ed. \textit{Welfare Reform and Faith-Based Organizations} (Baylor University Press, 1999), and the bibliographies kept at www.pewforum.org/faith-based-initiatives and at www.cpjustice.org/charitablechoice.
governance is the bureaucracy, in which all relations are (at least ideally) transparent in the strong sense. Without the transparency requirement, the state has available to it a variety of other mechanisms of government with different characteristics.

As Alasdair MacIntyre has pointed out, the bureaucratic institutions of the modern state have their own set of values, usually some *ad hoc* combination of utility and the respect for rights. Susan Mendus appends the observation that no matter which combination a particular bureaucracy institutes, it will express the view common to both conceptions (utility and rights). She calls this the ‘market’ view, a “particular, and disputed conception of autonomy” which is opposed to the development of ‘genuine’ autonomy because it disrupts a person’s allegiance to the smaller groups in which personal autonomy develops. Mendus and MacIntyre agree that this view of autonomy “is deeply at odds with the values constitutive of many genuine communities.” Needless to say, an integralist would share their concern about bureaucracies and the values it

---


50 MacIntyre, ibid., pp. 140-2.


52 Ibid. For a much more robust treatment of this problem, see Claude LeFort, *The Political Forms of Modern Society: Bureaucracy, Democracy, Totalitarianism* (MIT Press, 1986), chs. 3-4.
imposes, and would support efforts to integrate non-bureaucratic institutions into the mechanisms of government.\textsuperscript{53}

3.5 Conclusion

My argument in this chapter has been mostly negative: that the distinction between the public and the private spheres need not be explained in terms of autonomy as Rawls does, nor must it require that traditional societies and practices be made transparent, as Shklar does on at least some occasions. These negative arguments clear the way for the development of a liberalism that can accommodate the integralists, and that will allow a liberal regime to cooperate in its tasks of government with a wide variety of religious, traditional, and non-bureaucratic institutions.

If the argument of this chapter goes through, then it seems that liberalism would be more stable if it were not committed to autonomy. If we were to place Rawlsian political liberalism in a head to head comparison with a liberalism that captured a number of features of democratic cultures, was stable, but was not committed to autonomy, Rawls’s own arguments ought to lead him to choose the more stable of the two liberalisms.

Before we end Part One, we will look in the next chapter at an area where liberal political philosophy seems especially weak, namely, how to think about punishment. Punishment is, on its face at least, an area in which equality and autonomy are absent.

\textsuperscript{53} Larmore, \textit{Patterns of Moral Complexity}, pp. 41-2, admits the impersonal nature of bureaucracies, but sees this as a trade-off for greater transparency and predictability, which enable us to pursue a more predictable private life. He does not explain why predictability is an interest of a higher order than those trumped by bureaucracy.
The punisher and the punished are in a relationship of inequality, in which the former rules over the later. The citizens’ intuitions are conflicted about this, with some people putting more emphasis on the good of social order, and others focusing on the harm done to the prisoners. Liberal political philosophers, in their attempts to conceive of an ideal society in which all coercion is in some sense self-imposed, have had to avert their eyes from the practice of punishment, leaving it to utilitarians, libertarians, religious believers, and psychotherapists. But where there is inequality of power, there is a chance of abuse, inhumanity, and cruelty. There is also the possibility of mercy, the virtue that links power with humanity. If a state can punish well, it can increase stability not just in the present, but for future generations as well.
CHAPTER FOUR

LIBERAL AUTONOMY OR HUMANE PUNISHMENT?

The topic of punishment always invites discussion of mercy, since mercy connects the exercise of power with a concern for the moral dignity of human beings. In this chapter I hope to show four things: 1) that Rawlsian liberalism cannot reconcile its commitments to autonomy and equality with any regime’s need for a state-run system of criminal justice and punishment; 2) that any theory of the state has to have a theory of punishment as one of its central parts; 3) that attractive theories of punishment ought to balance respect for the humanity of the offender with the need for order and stability in society; 4) that punishment can profitably be seen as an extreme form of moral education, as some contemporary philosophers have urged. If these conclusions are right, then we have another reason to reject autonomy and tolerance in favor of benign heteronomy and mercy as the characteristic values of a stable and humane liberalism.

4.1 Liberalism & Punishment

One might think it quite evident that a theory of justice ought to include a theory of punishment. Plato, Aristotle, Bentham, Mill, Kant, and Hegel, to take major examples, all thought punishment was an essential part of a theory of justice. John Rawls, however, did not say much about punishment in *A Theory of Justice*, and he says nothing about punishment in his later writings. It is not even clear what a Rawlsian theory of
punishment would look like. In *A Theory of Justice* Rawls talks of punishment briefly in three places, but what he says there is underdeveloped at best, and seems to be just self-contradictory and incoherent. As a result, the vast literature inspired by Rawls’s *Theory* has little to say about punishment. In this chapter, I wish to show that Rawls’s own statements about punishment do not lead to a coherent theory, and that strict Kantian retributivism is the theory most consistent with the premises of justice as fairness. This is unfortunate for Rawls, since Kant’s theory of punishment is among the most unattractive theories around.

In chapters seven and nine, I will consider what an Anselmian view of punishment looks like, and how such a view would alter our understanding of the purposes of punishment, especially criminal punishment, from the theories of punishment currently fashionable. Rawls thought that a theory of punishment should only be determined after one develops principles for just institutions in a just society under ‘ideal’

1 Rawls discusses punishment in TJ, §§38-39, 48, and at the end of §86. Interestingly, his essay “Two Concepts of Rules” (*Collected Papers*, ch. 2) is considered a classic defense of a utilitarian theory of punishment, and is included in several anthologies of articles on punishment. But Rawls seems never to have returned to this topic at length.

2 A possible exception might be Robert Nozick, who made punishment a central issue in his argument against anarchy. See *Anarchy, State, and Utopia* (Basic Books, 1974) ch.4.

3 I use the word ‘fashionable’ deliberately, since it seems that attitudes towards punishment are as fickle as fashion trends, and brook as little opposition. Today, we are in a get-tough on crime trend, although this seems to have peaked. In the 1940s, on the other hand, rehabilitationist theories were so fashionable in the English-speaking world that even as popular a writer as C. S. Lewis could not find any journal in the U.S. or England that would publish his anti-rehabilitationist essay, “The Humanitarian Theory of Punishment”. Lewis was forced to publish it in an Australian journal, *Twentieth Century* (March 1949). It was reprinted in *God in the Dock* (Wm. B. Eerdmans, 1954).
conditions of ‘strict compliance,’ in which all citizens follow the law. Although Anselm, too, uses the device of an ideal society of strict compliance as the standard by which the current society should be judged, he nonetheless makes punishment a central feature of his view of politics and governance. For Anselm, punishment not only defends and deters, it also teaches, corrects, and persuades; and the latter qualities are the more important. He places a high priority on the rehabilitation and conversion of the offender, believing that this will lead to a more stable order than retributive and deterrent punishments, and one that more adequately respects the offender’s dignity. Mercy is central to this view of punishment, and it is by considering the role of mercy in punishment that we will best be able to understand its role in developing a broader theory of the just society, one characterized by benign heteronomy.

4.2 Rawls on Punishment

It is not clear what a Rawlsian theory of punishment would look like. In *A Theory of Justice*, Rawls insists that full-fledged theories of punishment are not part of ideal theory, because *ex hypothesi*, everybody in a well-ordered society has internalized the principles of justice, has allegiance to the just regime, and acts from the principles of justice in pursuing the good. In effect, Rawls assumes that in a well-ordered society, everybody will comply with the law—it is for this reason that he labels his ideal theory as a ‘strict-compliance theory’, as distinguished from a ‘partial compliance theory.’ At a later point, Rawls uses this distinction between strict and partial compliance theory to claim that criminal justice is different in kind from distributive justice: criminal justice is part of ‘partial compliance theory’ (a part of non-ideal theory), while distributive justice
is part of “strict compliance theory, and so part of the ideal scheme.” Rawls makes this
distinction in the context of wondering whether there will be any need for law and penal
sanctions in the ideal scheme with its hypothesis of strict compliance. He concludes that
even under these ideal circumstances, the principle of liberty would require that the
liberal society be governed by the rule of law, and law is not law if there are not some
penalties for noncompliance. Rawls calls this ‘Hobbes’s thesis’:  

> Even under reasonably ideal conditions, it is hard to imagine, for example, a successful income tax scheme on a voluntary basis. Such an arrangement is unstable. The role of an authorized public interpretation of rules supported by collective sanctions is precisely to overcome this instability. By enforcing a public system of penalties, government removes the grounds for thinking that others are not complying with the rules. For this reason alone, a coercive sovereign is presumably always necessary, even though in a well-ordered society sanctions are not severe and may never need to be imposed. Rather, the existence of effective penal machinery serves as man’s security to one another. . . . It is clear from the preceding remarks that we need an account of penal sanctions however limited even for ideal theory. Given the normal conditions of human life, some such arrangements are necessary.

Rawls wants to emphasize that penal sanctions are necessary, even in this ideal ‘society
of angels,’ because liberty demands them. Appealing to Hobbes’s Thesis allows Rawls to
make two points: i) since laws and penalties are necessary even in the ideal scheme, they
will certainly be necessary in the nonideal reality; ii) because in the ideal scheme the
penalties are justified by an appeal to liberty, the same justification should hold in the
nonideal situation. Rawls argues that liberty, rather than retribution or denunciation, is

---

4 TJ, §48, p. 277; for strict compliance theory, see §25, pp. 125-6.

5 Ibid., §38, pp. 211-2
the justification of punishment. In doing so, he was deliberately staking out a position in the lively debates about punishment of his day, aligning himself with H.L.A. Hart against both K. G. Armstrong and Joel Feinberg. A quick summary of these debates will help us follow the arguments in the rest of the chapter.

4.3 A Short History of Punishment Theories

Historically, the practice of punishing has been justified in three major ways: by Christianity as somehow redemptive, by utilitarians as a deterrent, and by Kant (following Hammurabi and his descendants) as being somehow retributive. By the middle decades of last century, though, the utilitarians held the upper hand. Influential articles by J. D. Mabbot, Anthony M. Quinton, and Rawls himself showed how the standard charge against utilitarianism—that it allowed for punishing innocent people if that would contribute to the common good—could be neatly evaded. Following a suggestion by Sir David Ross, they argued that the justification of punishment should be divided into two stages: the justification of the institution of punishment (the law with its penalties, as well as the entire criminal justice system) and the justification of the particular act of punishing. If the general practice of punishment could be justified, then the particular acts

---

6 Ibid., p. 212.
9 Rawls, “Two Concepts.”
10 W. D. Ross, The Right and the Good (Oxford University Press, 1930) pp. 56-64, esp. 61-2.
did not require an independent justification, since they were mere applications of the justified practice. (This is a common strategy of rule utilitarianism versus act utilitarianism.) Since it does not seem that one can justify an institution or rule whose purpose is to punish innocent people when the common good so demands, there will not be a law creating such an institution, and anyone who does punish an innocent person will be breaking the law and herself ought to be punished.

With this defense in hand, the utilitarians made an alliance with progressive Christians against the retributivists. Religiously inspired movements for penal reform tended to emphasize the need for society to show mercy, to care for the soul of the prisoners, to help them repent and reform their lives. The utilitarians argued that if these strategies could prevent recidivism, then they could lead to the greatest happiness at the least cost, just as Jeremy Bentham’s principles of punishment required. Both the religious ‘reformatory’ view of punishment and the utilitarian ‘deterrence’ view were forward-looking views of punishment, while the retributive view was backward-looking. Together, these views had the retributivists on their heels. K. G. Armstrong provides a nice summary of the utilitarian arguments against retributivism:

Retributive punishment is only a polite name for revenge; it is vindictive, inhumane, barbarous, and immoral. Such an infliction of pain-for-pain’s sake harms the person who suffers the pain, the person who inflicts it, and the society which permits it; everybody loses, which brings out its essential pointlessness. … By making the punishment of wrongdoers a moral duty, the retributive theory

---


removes the possibility of mercy. … In any case, the theory is not only morally indefensible but completely inadequate in practice to determine what penalty the criminal should suffer in each case.

This last point was often considered the most powerful. Retribution requires that somehow the punishment should fit the crime. But it often will be difficult to find punishments that fit every crime: what’s the equal and opposite reaction to bigamy? Forgery? The rape of your innocent daughter? The murder of a promising young doctor in the bloom of youth by a desperate junkie with two months to live? Where utilitarianism promised rational precision based on modern social science research, retributivism seemed wrapped in primitive mystery and symbolism, and was incapable of being put into effect.

In an article providing a classic defense of retributivism, Armstrong detaches retributivism from Kant’s strict *lex talionis* requirement (about which more below). Retributive considerations need only set an upper limit to punishment, while other considerations (including those of mercy, excessive cost to society, or diminished responsibility) can mitigate that punishment. Rendered this way, Armstrong argues, only retributivism truly respects the criminal as a moral agent: “If we penalize a criminal according to what he has done, we at least treat him like a man, like a responsible moral agent. If we fix the penalty on a deterrent principle…we are using him as a means to somebody else’s end, and surely Kant was right when he objected to that!” So much for deterrence. Furthermore, citing C.S. Lewis’s influential criticism of “The Humanitarian

---

13 Ibid., p. 152.
Theory of Punishment," he argues that purely forward-looking reformative theories would require that no matter how minor the crime actually committed, the prisoner should not be released until the experts decided he was completely cured of all criminality, i.e., that not only had he been cured of the desire to commit the crime he actually committed, but that he was cured of the desire to commit crime altogether. The curing of all criminal inclination might take substantially longer than the ordinary sentence for the crime actually committed, but the sentences would be justified if the reformative theories were. Moreover, Armstrong argues, “since prevention is better than cure, why wait until he commits a crime?”:  

On the Reformatory theory of penalty fixing, it is the tendency to commit crimes that we want to eliminate, so if a man has the tendency let him be penalized before the damage is done. Let him be penalized for what he is, not for what he does, and let him be made over into what the authorities (or their experts) want him to be.

So much for reformation. Armstrong’s article did not give a positive argument for retributivism, but simply criticized the other two views so that retributivism looked like the only theory that protected the individual against the totalitarian state. These criticisms set the agenda for H.L.A. Hart’s defense of a consequentialist approach to punishment, to which Rawls signs on in *A Theory of Justice*.

---

14 Ibid.

Joel Feinberg developed another view of punishment somewhat different from the ones above, what Rawls (following Hart) calls the ‘denunciatory theory’ of punishment. This view holds that punishment is distinguished from mere penalties (e.g., “parking tickets, offside penalties, sackings, flunkings, and disqualifications”) by a characteristic ‘expressive function’.

Punishment is a conventional device for the expression of attitudes of resentment and indignation, and of judgments of disapproval and reprobation, on the part either of the punishing authority himself or of those ‘in whose name’ the punishment is inflicted. Punishment in short has a *symbolic significance* largely missing from other sorts of penalties.

---


19 Ibid., p. 98. Feinberg lists six common functions of punishment that depend on its expressive function (p. 101ff):

1) Deterrence: it deters a criminal to know that action A is condemned.
2) Authoritative disavowal: US plane fires on Soviet, Soviets demand that pilot be punished, or else it expresses that the pilot acted as an agent of the US. A corporation punishes an employee to disavow his acts.
3) Symbolic Acquiescence: I want action A punished because I want the state, in the name of the people, to condemn A in a way that makes it stick.
4) Vindication of the law: merely passing a law that is not enforced expresses that we don’t really care about the values in the law. Punishment makes it clear that we do.
5) Punitive damages in civil suits: Here’s what you owe the plaintiff, and here’s what you owe society, you jerk.
6) Absolution of others: If you punish Whittaker Chambers, you absolve Alger Hiss, and vice versa. If you say John raped Mary you preserve Mary’s honor, while if you say Mary consented, you absolve John.
Feinberg thinks that this expressive function must be a part of any definition of punishment, and that a sound retributive theory ought to try to make the ‘condemnatory aspect’ of the punishment (rather than the pain) fit the crime. He thinks that we have developed physical hard treatment simply as the conventional way for our society to express its condemnation, but he is ambivalent about this, and seems to think that it might be possible to punish without pain.

Hart subjects this theory to what is often considered a devastating critique. First, he argues, denunciation lets the judge and the community off the hook. For while it appeals to a judge’s vanity to invite him to act as ‘the mouthpiece of the moral sentiments of society,’ his job is really the much more difficult one of determining facts about a particular case and thinking about the effects of what the society is doing. And while it reassures a community to have its scale of moral evaluations confirmed from the bench, it often happens that in the course of applying justice to a particular case, the decisions of the judge serve to challenge the community’s assumptions about what is right and wrong. It would be a loss for the law to lose this capability. Secondly, he argues, a pluralistic society will often have many contradictory reasons for condemning a certain action, and often not agree that a particular action should be condemned, so if the judge is asked to speak for the community he will often end up expressing only the views of his social class. Finally, he argues, if denunciation is to be combined with retribution, so that

20 Hart, *Punishment and Responsibility*, pp. 170-173. It might be the case that Feinberg’s theory is not susceptible to Hart’s criticisms, if one reads it not as an attempt to *justify* punishment so much as to examine one among the many functions of punishment. In any case, Hart cites a number of people who do think denunciation justifies punishment, so he is certainly not attacking a straw man.
sentencing is a “search for a penalty the severity of which, like a gesture, will aptly express a specific feeling of revulsion or moral indignation,” this seems to be too vague to act as a justification for anything, “a semi-aesthetic idea which has wandered into the theory of punishment.” 21 We do not punish in order to condemn, although we might condemn while punishing; to get this backwards is to “subordinate what is primary to what is ancillary.”22

Hart’s own theory of punishment was, quite clearly, the most sophisticated theory at the time Rawls was writing A Theory of Justice. It is sometimes called a mixed theory, because it identifies the goal of the practice of punishment as the reduction of crime, or more generally to the protection and promotion of the common good, as the utilitarians do, while agreeing with the retributivists that questions about who may be punished and in what amount should be determined by considerations of guilt and culpability rather than deterrence. Hart calls the first stage “the general justifying aim of punishment” and the second the question of “the distribution of punishment.”23 He holds that theories about the distribution of punishment can be developed independently of theories about the justifying aim of punishment. It is not an argument in favor of retribution in general to point out that we are all retributivists when it comes to the distributive question of

21 Ibid., p. 172.

22 Ibid.

23 We have already seen that Rawls and Quinton defend versions of this distinction well before Hart, but it is Hart’s terminology and arguments that became canonical. Rawls and Quinton argued that the distinction saves utilitarianism from the criticism that it would countenance punishing the innocent, since utilitarians could never justify an institution that had harming the innocent as an aim. Hart accepts and extends their arguments, as we shall see.
choosing whom to punish. A deterrence theorist can agree that it would be immoral to punish the innocent, even if it would seem to have advantages for crime prevention, and yet still think that the prevention of crime, and not restoring society to the *status quo ante*, is the aim of criminal justice system. On the other hand, only the utilitarian has the option of a mixed theory—someone who thinks retribution is the general justification of punishment would also be a retributivist about the distribution of punishment.

Hart was a sort of consequentialist, and had little patience for retributivist accounts of the aim of punishment. But he argues at great length to defend the traditional legal doctrine of *mens rea*, i.e., that an offender’s intentions and degree of responsibility are relevant to his guilt and to what penalty is appropriate. England in the 1950s and early 1960s was quite taken with psychiatry and psychological determinism, such that many people took it quite seriously when the first detailed psychiatric studies suggested that many of those who committed crimes were mentally ill, and so were not responsible for their actions in the traditional sense.\(^{24}\) Some even suggested that the law abandon the idea of responsibility altogether and move to a two stage system: first, strict liability\(^{25}\) in sentencing, after which state-appointed psychiatrists would determine whether those who broke the law did so responsibly or whether they needed psychiatric treatment before

\(^{24}\) Cf. Barbara Wootton, *Social Science and Social Pathology* (George Allen and Unwin, 1959) esp. ch. VIII.

\(^{25}\) In cases of strict liability, “it is no defense that the accused did not offend intentionally, or through negligence” (Hart, *Punishment and Responsibility*, p. 20). Strict liability offences include parking tickets, laws against selling liquor to an intoxicated person, and laws against possessing an altered passport.
they can be released.\textsuperscript{26} Since psychiatry had contributed to skepticism about the idea of ‘guilt,’ most people assumed it was also calling into question the doctrine of \textit{mens rea} (literally ‘a guilty mind’). Hart thought this proposal serious enough that he felt he needed to provide a foundation for the doctrine of responsibility that did not depend on the idea of guilt. It was this proposal that Rawls suggests can justify the existence of legal sanctions even in a well-ordered society, although, we will suggest, Hart’s view is ultimately unavailable to Rawls.

Hart thinks that he can “provide a rationale for most of the existing excuses which the law admits in its doctrine of \textit{mens rea}” by appealing to the ideas of fairness, justice, and the value of individual liberty.\textsuperscript{27} What would we lose, he asks, by moving to a system where all liability was strict? For sure, we would lose the ability to predict the outcome of our decisions, to decide whether we will make ourselves eligible for punishment by the state or not. If we could be punished for actions we didn’t choose to do, then we could not be sure that we wouldn’t be punished tomorrow. That would make it more difficult to plan our lives—we would not be able to enter into a contract, let alone a marriage, because we couldn’t promise that some action we did by mistake or accident wouldn’t land us in jail. Such a system might protect society from criminals, but at great cost to freedom.

Hart argues that a system of strict liability ignores a distinction that is essential to human society. People are not, as the strict liability system would seem to assume, simply

\textsuperscript{26} Cf. Wooton, \textit{Social Science and Social Pathology}, pp. 334-7

\textsuperscript{27} Hart, \textit{Punishment and Responsibility}, p. 181.
“so many bodies moving in ways which are sometimes harmful and have to be prevented or altered.” Instead, 28

persons interpret each other’s movements as manifestations of intention and choices, and these subjective factors are often more important to their social relations than the movements [themselves]. If one person hits another, the person struck does not think of the other as just a cause of pain to him; for it is of crucial importance to him whether a blow was deliberate or involuntary. … If you strike me, the judgment that the blow was deliberate will elicit fear, indignation, anger, resentment: these are not voluntary responses; but the same judgment will enter into deliberations about my future voluntary conduct towards you and will color all my social relations with you. Shall I be your friend or enemy? Offer soothing words? Or return the blow? All this will be different if the blow is not voluntary. This is how human nature in human society actually is and as yet we have no power to alter it.

Hart thinks that these distinctions are so basic to our social life that the law cannot ignore them. People are not things; they have volition, and therefore the law should acknowledge they are responsible for their actions. By waiting to interfere until a crime has been committed and is proved to have been committed voluntarily, our system of punishment recognizes “that a man’s fate should depend upon his choice and this is to foster the prime social virtue of self-restraint.” 29 We are unwilling to punish preemptively, because we believe that social trust depends on a defeasible presumption that each person desires not to injure us and has the self-restraint to act accordingly. To hold otherwise, that men do not have the ability of self-restraint, or that they are not

28 Ibid., pp. 182-3. Emphasis original.

29 Ibid., p. 182.
averse to injuring us, would fail to reflect the assumptions that “pervade the whole of our social life.”

4.4 Does Rawls Have a Theory of Punishment?

Rawls appeals to exactly this passage in the context of his argument that even strict compliance theory needs an account of law and punishment. The principle of liberty, he claims, leads to the principle of responsibility, without needing to appeal to retribution or denunciation. If we admit that citizens are rational, and we give the “appropriate weight” to liberty, then it follows that “unless citizens are able to know what the law is and are given a fair opportunity to take its directives into account, penal sanctions should not apply to them.” He makes no argument for this, but instead cites Hart’s discussion, which we have just been summarizing.

That Rawls simply signs on to Hart’s analysis should strike us as quite curious. Hart is a utilitarian, and Rawls is not. As a social contract theorist, Rawls starts with the individual and constructs society, while Hart’s philosophizing always starts by considering society and its actual practices. What’s more, Hart does not share Rawls’s presupposition that natural assets are irrelevant to questions of justice, or at least, he does not share the corollary that all men are equally capable of following the law.

---

30 Ibid., p. 183.
31 TJ §38, p. 212n25.
32 Ibid., p. 212.
Hart argues that our responsibility for our actions is something that admits of degrees, because our capabilities for acting otherwise admit of degrees. In particular, he regards self-restraint as a “prime social virtue” that the law should foster, but which is not natural or equally distributed among all people. In our ordinary morality, our assessments of a person’s ability to control her actions are “of crucial importance,” and the law ought to respect this. Indeed, the “important general principle” animating Hart’s discussion amounts to the claim that considerations of moral ability—i.e., whether the person who injured another a) possessed and b) exercised the virtue of self-restraint or not, and if not, whether c) he was given the fair opportunity to acquire and/or exercise this virtue—are necessary for justifying the doctrine of responsibility.

We can use Jane Austin’s *Emma* to illustrate this idea of virtue as a moral ability that varies in degrees. A person who is from a socially advantaged position ought to have many social virtues, and so ought to be capable of greater self-control than someone who lacks social virtues. Mr. Knightley is therefore right to criticize Emma Woodhouse for a minor slip in etiquette, because she had the ability to conform to the rules of etiquette, and deliberately chose not to. Her mildly critical joke in the wrong context, which would

---


34 Ibid.: “Underlying these separate points there is I think a more important general principle.”

35 Though not sufficient—Hart is clear that the reason we are concerned about moral ability is for other reasons of social life, and not simply because the moral virtue of self-restraint is the *summum bonum* or is even good per se. Rawls agrees that the presence or absence of the virtue of self-restraint is not sufficient to justify the doctrine of responsibility, but he disagrees with Hart that it is necessary or even relevant to such justification. See TJ, §48 for his argument against moral desert.
be morally insignificant for another, is for Emma a “manifestation of intention and choices,” to use Hart’s phrase, for which she can be faulted. On the other hand, Emma is right to refrain from criticizing Harriet Smith for her relative lack of social graces—she has not been brought up with the necessary refinement, and so is not always able to recognize how she ought to behave. Emma is also right to be critical of Mrs. Elton for lacking the social graces her privileged background gave her the opportunity to acquire.

As Hart notes, our ordinary social life pays much attention to such considerations. Yet Rawls, famously, declares that “natural assets and the contingencies of their growth and nurture in early life are arbitrary from a moral point of view,” and that justice as fairness rejects the idea of moral desert. Hart’s idea of responsibility is more complicated than Rawls can allow into his ideal theory, because Hart’s theory is founded on (non-ideal) social realities that Rawls has to set aside in order to motivate the original position. It seems, therefore, that Rawls has set himself against the principle upon which Hart founds his view. If we don’t allow Rawls to appeal to Hart’s analysis, though, it is not clear how he can ground the principle of responsibility, or the practice of punishment under conditions of partial compliance. His strategy was supposed to be that he would give a theory of punishment for conditions of strict compliance, and then use the philosophical principles developed at that stage to justify (and even reform) the practice

36 TJ, §48, p. 274. Note, this idea of Rawls’s is quite fundamental—he says that his views of moral personality and the original position are based on the assumption that natural assets and their development are irrelevant to the concerns of ideal theory. See TJ, §3, p. 11.

37 Ibid., p. 273.
of punishment in real life. But he says so little about punishment, and what he says is so inadequate, that he doesn’t seem to have an account of punishment at all.

If we look at a second set of comments Rawls makes about punishment, we are left even more confused. Rawls announces that the purpose of the criminal law is to uphold basic natural duties, i.e., those duties that would be chosen by the parties in the original position before the erection of any institutions or social practices. Rawls lists the following as natural duties:

- The duty of helping another in need when it does not involve excessive risk to oneself;
- The duty not to harm or injure another;
- The duty not to cause unnecessary suffering;
- The duty to support and comply with just institutions and to establish such institutions where they do not already exist;
- The duty not to deprive another of life, liberty or property; and
- The duty to respect other human beings as possessors of a sense of justice and a conception of the good.

He says that violations of natural duties should be criminalized; moreover, he claims that a propensity to violate these duties is “a mark of bad character” that makes one worthy of retributive punishment based at least in part on moral desert. In defense of this last claim, Rawls cites Feinberg’s article defending the view that punishment is essentially denunciatory, even though he has said earlier that punishment should not be justified by

---

38 Ibid., p. 276.
39 Ibid., §19.
40 For all but the last natural duty, see ibid. For the last, see §51.
41 Ibid., §48, pp. 276-277.
This appeal to Feinberg’s article is quite puzzling, given Hart’s criticisms of the denunciatory justification of punishment, and Rawls’s attempt to enlist Hart on his side.

In his last set of remarks on punishment, in the context of a thought experiment about people in a well-ordered society who find that the sense of justice is not a good for them, he seems to reverse himself again. He argues that these people should be punished if they act on their view of the good, because the stability of society would be undermined if they weren’t. He repeats his claim that the practice of punishment is grounded on the need for equal liberty and the rule of law, and then goes to make a curious remark:

Those who find that being disposed to act justly is not a good for them cannot deny these contentions. It is, of course, true that in their case just arrangements do not fully answer to their nature, and therefore, other things being equal, they will be less happy than they would be if they could affirm their sense of justice [to be valuable]. But here one can only say: their nature is their misfortune.

This comes perilously close to arguing that one’s moral nature is not ‘arbitrary from the moral point of view,’ that fortune is relevant to matters of justice, and that those who are unlucky might need to be punished so that society will be more stable. It is not clear how punishing (or threatening to punish) the misfortunate for their unfortunate natures is consistent with the doctrine of mens rea. What is clear is that if (somehow) there were

---

42 Ibid., p. 277n41; cf. §38, p. 212. See Feinberg, “The Expressive Function of Punishment.”

43 Ibid., §86, pp. 503-505.

44 Ibid., p. 504.
criminals in the well-ordered society, punishment and other legal penalties should be designed to uphold the principles of justice as fairness.

We have seen that Rawls’s diverse claims about punishment are, uncharacteristically for Rawls, unsystematic and terrifically unclear. Nonetheless, it seems fair to attribute to Rawls the position that the general justifying aim of punishment is to encourage and enforce compliance (strict compliance in the ideal theory) with the principles underlying the social contract. Since Rawls never returned to these questions, we will have to look for help in the secondary literature, beginning with Jeffrie G. Murphy’s argument, published just after A Theory of Justice came out, that Rawls can and should adopt a Kantian theory of punishment.45

4.5 Punishment in Murphy’s Rawlsian Liberalism

Kant held that to be punished justly, a criminal must be punished voluntarily. This seems a difficult requirement to meet, since most criminals seem quite unwilling to undergo punishment, especially capital punishment. Murphy realized that the original position and other elements of Rawls’s theory provided a rigorous and clear way to make sense of this claim that coercion is compatible with autonomy. On Murphy’s reconstruction, Kant’s position is as following: X can compel Y to do A, even if Y does not want to do A, so long as Y could have rationally willed to do A. In that case, Y’s desire not to do A is irrational. So, e.g., the coercion of a debtor “is entirely compatible

with everyone’s freedom, including the freedom of the debtor, in accordance with universal laws.”\textsuperscript{46} Were the debtor in something like Rawls’s original position, she would choose laws that allowed for the coercion of debtors, knowing that when the veil is lifted some of her own desires might be thwarted. Therefore, when she is coerced, it can be said that she rationally wills it, even though her irrational desires might go the other way. In this way, coercion does not violate her freedom.

Kant’s contractarian view of the state leads to his strong retributivist view of punishment. In a passage that could have come right out of Rawls, Kant defines justice as simply “the aggregate of those conditions under which the will of one person can be conjoined with the will of another in accordance with a universal law of freedom.”\textsuperscript{47} Every crime is a violation of the social contract, of the mutual obligations that reason shows we have towards each other, and justice requires that the network of mutual obligations be completely restored. Here’s how Murphy summarizes the key Kantian points:\textsuperscript{48}

If the law is to remain just, it is important to guarantee that those who disobey it will not gain an unfair advantage over those who do obey voluntarily. It is important that no man profit from his own criminal wrongdoing, and a certain kind of ‘profit’ (i.e., not bearing the burden of self-restraint) is intrinsic to criminal wrongdoing. Criminal punishment, then, has as its object the

\textsuperscript{46} Ibid., p. 52. The quote is from part I of Kant’s \textit{The Metaphysics of Morals} (p. 232 in the Prussian Academy edition). Emphasis is Murphy’s.

\textsuperscript{47} Quoted in ibid., p. 54. Cf. Kant, MM p. 230.

\textsuperscript{48} Ibid., pp. 54-5. Cf. Kant, MM, 331-2. It should be noted that Murphy later considered this view to be more Rawlsian than Kantian, since he later came to have doubts as a Kant scholar that Kant had a coherent theory of punishment at all. See “Does Kant Have a Theory of Punishment?” \textit{Columbia Law Review} 87 (1987) pp. 509-32.
restoration of a proper balance between benefit and obedience. The criminal himself has no complaint, because he has rationally consented to or willed his own punishment. That is, those very rules which he has broken work, when they are obeyed by others, to his own advantage as a citizen. He would have chosen such rules for himself and others in the original position of choice. And, since he derives and voluntarily accepts benefits from their operation, he owes his own obedience as a debt to his fellow-citizens for their sacrifices in maintaining them. If he chooses not to sacrifice by exercising self-restraint and obedience, this is tantamount to his choosing to sacrifice in another way—namely, by paying the prescribed penalty:

(Quoting Kant) ‘A transgression of the public law that makes him who commits it unfit to be a citizen is called… a crime…. What kind and what degree of punishment does public legal justice adopt as its principle and standard? None other than the principle of equality… that is, the principle of not treating one side more favorably than the other. Accordingly, any undeserved evil that you inflict on someone else among the people is one you do to yourself. If you vilify him, you vilify yourself; if you steal from him, you steal from yourself; if you kill him, you kill yourself… To say, “I will to be punished if I murder someone” can mean nothing more than “I submit myself along with everyone else to those laws which, if there are any criminals among the people, will naturally include penal laws.”

This analysis of punishment regards it as a debt owed to the law-abiding members of one’s community; and, once paid, it allows re-entry into the community of good citizens on equal status.

Murphy holds that a criminal wills punishment because he is a free and equal rational being, even though the criminal can assure us that he does not want to be punished.

Murphy acknowledges that this has a paternalistic feel to it, as though the state is telling the criminal, “We know what you really want more than you do.” In real life, this argument can be susceptible to abuse, Murphy notes (e.g., the torturer who dabbles in Kant tells the tortured, ‘you would will this, were you only more rational’). Nonetheless, Kant insists that freedom and autonomy require that we will rationally; irrational desires only enslave. From the original position we can see what is the rational thing to will, and
in social contract theory, we do will it. The device of the original position blocks the ‘Kantian torturer’ objection—rational agents would not choose torture or similar practices of punishment behind the veil of ignorance.

And rational agents would choose a system of strict retribution behind the veil of ignorance, at least on Kant’s account. Since the contract situation requires that all citizens be treated equally, anyone who gains an unfair advantage from the contract needs to be brought back to a state of equality, so that the balanced situation of mutual advantage is restored. This is the free-rider model of criminality.\textsuperscript{49} Since the social contract situation best guarantees personal autonomy and freedom, the form of punishment selected by this system would also best guarantee personal autonomy and freedom. Hence we get Kant’s at first startling claim that:\textsuperscript{50}

Even if a civil society were to be dissolved by the consent of all its members (e.g., if a people inhabiting an island decided to separate and disperse throughout the world), the last murderer remaining in prison would first have to be executed, so that each has done to him what his deeds deserve and blood guilt does not cling to the people for not having insisted on his punishment; for otherwise the people can be regarded as collaborators in this [the murderer’s] public violation of judgment.

Though counterintuitive, this claim makes sense, if we grant the social contract theory and Kant’s additional premise that “there is no similarity between life, however wretched


\textsuperscript{50} Kant, MM, p. 333.
it may be, and death, hence no likeness between the crime and the retribution unless death is judicially carried out upon the wrongdoer.”

Murphy himself argues that the Kantian/Rawlsian theory is a good theory of punishment, but only if a society in general meets the standards of liberal justice that Kant and Rawls describe. The theory meets the formal standards of justice because it treats the criminals as free and equal beings, but were such practices of punishment established in a society with great inequality, strict retribution would be materially unjust. To use Rawlsian jargon, Murphy’s view is that this theory is good for ideal theory, but not for non-ideal theory, except as a guide for reform.

If Murphy is right that this strict retributive theory is a reasonable consequence of the Kantian premises and methods of *A Theory of Justice*, many liberals would conclude that this is a serious, perhaps fatal weakness to that project. For starters, it would undoubtedly be unstable on a number of grounds. It often will be difficult to find punishments that fit every crime, and yet failure to do so would lead to widespread injustices according to the theory, and would put strains on a people’s commitment to the regime. Then too, there is the problem that Martha Nussbaum credits Seneca with

51 Ibid. Emphasis original.

52 See Murphy, “Marxism and Retribution,” pp. 57ff.


articulating: since in every society circumstances corrupt man very deeply and from a very early age, even a properly equitable sense of justice, combined with even the finest tuned sense of retribution (perfectly balancing eye for eye), will end up leading man to hate almost everybody to some degree. “A person who notes and reacts to every injustice, and who becomes preoccupied with assigning just punishments, becomes, in the end, oddly similar to the raging ungentle people against whom he reacts. Retributive anger hardens the spirit, turning it against humanity.” Making a similar point, Jean Hampton notes that it is because retribution requires that evil be met with an equal evil (rather than, say, with some good), that it strikes so many people as “too close to revenge.”

In addition to the problem of stability, the strict retributivist view of punishment seems to violate our ordinary intuitions about justice. Questions of mitigating personal history or aggravating intention could not be considered either in determining guilt or in fixing a sentence. A retarded murderer who was abused as a child and, as member of a minority group, faced severe discrimination is not to be treated any differently than a professional killer with a military background.

Andrew von Hirsch points out that this retributivist view has some deep theoretical problems when it comes to explaining morally intuitive differences in sentencing. Von Hirsch notes how odd it is to say that penalties should be set by the


amount of freedom a criminal unfairly claims for himself. “It is…obscure to say the robber deserves more punishment than the burglar because, somehow, he has arrogated to himself a greater degree of unwarranted freedom vis-à-vis others than the burglar has.”

And it is equally odd to describe “victimizing crimes, such as armed robbery, in terms of the advantage a robber gains over uninvolved third parties, rather than in terms of the intrusion into the rights of actual or potential victims.” Von Hirsch notes that tax evasion is a good example of a crime in which one takes more than one’s fair share without hurting anyone in particular. Yet it is just weird to think that tax evasion is the paradigmatic case by which penalties for murder and rape should be set.

As we noted earlier, Murphy developed this Kantian theory of punishment by using relevant elements of Rawls’s *A Theory of Justice*. One might wonder whether Rawls’s political turn allows him to avoid these objections. Might not restricting the social contract to matters of constitutional essentials help avoid the more stringent requirements of full-fledged Kantian retributivism? It does not seem so, since the justification of punishment would seem to be one of those constitutional essentials upon which reasonable citizens should agree. Rawls’s liberal principle of legitimacy requires that the state’s exercise of power must be in accord with a constitution acceptable to reasonable people. Presumably a constitution would allow the regime to penalize or punish those who violated its principles, and so the power to punish would have to be a constitutional essential. (Even, the most essential ‘constitutional essential’?)

58 Ibid., p. 268.

59 Ibid., p. 269.
Indeed, it would seem that the political turn in Rawls’s thought makes it even harder to develop a coherent theory of punishment. It would not be a problem that there are numerous opposing justifications for the practice of punishment, so long as the different justifications of the practice arrived at similar decisions about what penalties to assess, or at least on the procedure for matching punishments and crimes. Unfortunately, while there does seem to be a consensus that we should punish only those who are guilty of a crime, there are deep disagreements about how much to punish them, or whether a lack of responsibility for one’s action should mitigate the penalty, or whether rehabiliting prisoners should be a goal of punishment, or whether the exercise of mercy is permissible or just. On top of this problem, some of the disagreements seem to stem from commitments that are particular to comprehensive doctrines: some religious people consider rehabilitation to be an important goal of punishment, while other religious people think retribution and strict justice to be more important. Many atheists and determinists think crime prevention is the obvious goal of punishment, and therefore that prisons should be unpleasant places to be, while those who think no soul is beyond redemption want prisons to be as humane as possible. It seems fair to conclude that punishment is an irreducibly moral question, one that is not susceptible to Rawlsian politicization.

In an article written over a decade after he laid out the Kantian retributivist theory of punishment, Murphy took up these questions about how to relate a theory of
punishment to a theory of the state. It will be helpful to examine his arguments before we turn to our ‘Anselmian’ solution to this problem.

4.6 Deterrence and the Liberal State

Murphy eventually came to reject the Rawlsian/Kantian theory of punishment, because he was persuaded by Robert Nozick’s criticisms of its underlying concept of ‘fairness.’ As a result, he came to embrace a libertarian social contract view of the state as the best framework from which to think about punishment. He suggests that because punishment is an evil, a radically intrusive restriction on freedom, “it is natural that persons committed to the liberal views of individual rights and a free society would, on both moral and political grounds, accept a system of criminal punishment only with great reluctance and as a last resort.” He adopts the two-part “strict scrutiny” test from constitutional law jurisprudence as a helpful guide to deciding whether the state should

60 Murphy, “Retributivism, Moral Education, and the Liberal State.”

61 Ibid., p. 10n11. Cf. Nozick, Anarchy, State, and Utopia, pp. 90ff; Rawls, TJ, §18. The principle of fairness holds that those who voluntarily benefit from the activities of a just institution are obligated to support that institution, whether or not they consent to the activities. Nozick objects to the idea that if other people can start an activity that benefits them and also benefits me, it imposes an obligation upon me to support that activity, whether or not I consent to it. In addition, he argues that even if there were such an obligation, it would not give the others the right to enforce it, any more than giving me a book which I did not ask for gives you a right to seize payment from me. There is, of course, an extensive literature about these competing visions of social contract theory. It is perhaps noteworthy that the principle of fairness does not appear in PL, although Rawls does respond to other criticisms by Nozick (cf. JAF, pp. 32, 51-52, 53, 83, 97). As Murphy notes, something like the principle of fairness is essential to the Kantian view of punishment.

punish in a given case. That test asks two questions: a) Is the goal in punishing a compelling state interest? b) Of all the ways to achieve this goal, is punishment the least restrictive alternative? He points out that by employing a Nozick-style social contract model he can reformulate the first question to ask: Would people living freely apart from society come together and exchange their liberties for the purposes of punishment? (Obviously, a reason would count as a compelling interest of the state if people found it compelling enough in the state of nature to form a state.) Social defense passes this test, Murphy plausibly argues, but retribution doesn’t.

Retribution fails in large part because this version of the social contract chooses a starting point that stacks the deck in favor of forward-looking theories: the whole point of coming together in a state is to prevent future violations of natural rights and freedoms. Insofar as retribution is backwards looking, seeking to restore what has been lost, it cannot provide a reason for giving up current freedoms for guarantees of future protection. Any argument that does pass this test will be some sort of deterrence or crime prevention theory, though retributive goals might serve the secondary purpose of placing some constraints on society’s need for security. If it is a secondary goal, then it might

\[63\] Ibid., p. 5.

\[64\] Ibid. “The obvious initial answer (one that gives comfort to deterrence and incapacitation theories of punishment) is self-protection—protection of these persons from outside threats (national defense) and from inside threats from the violent and non-cooperative persons in their midst (police power. … At least this one goal—definitive of even a minimal state—will surely strike rational persons as compelling and as clearly sufficient (if used only when necessary) to justify the curtailments of liberty involved.”

\[65\] Ibid., p. 6.
count as a rational state interest, but not a compelling state interest, and the demands of
deterrence would override when the two goals conflict.

4.7 Murphy’s Dilemma: Liberalism or Punishment?

So Murphy’s suggestion is that if we adopt a liberal theory of the state, the goal
and general justification of punishment should be the prevention of crime, and retributive
considerations should come into consideration only when we decide who to punish and
how much. Murphy’s strict scrutiny test leads him to hold that deterrence needs to be the
general justifying aim of punishment in a liberal state. But he also feels attracted to
retributive and other theories of punishment as moral theories, enough that he can write,
“There are times when I am tempted to drop liberalism in order to embrace [those
theories].” This leads him to conclude that “when moral accounts of punishment are
brought together with issues in social and political philosophy (including the philosophy
of constitutional law), we may not be able to retain everything that appeals to us in one
neat conceptual package, in one coherent system. Some fundamental theoretical choices
may be required.”

Murphy’s observation is accurate, and quite troublesome. Current liberal thought
does tend to develop its theories about the appropriate political regime in isolation from
moral reflection on punishment, and reverse is true for contemporary philosophical
reflection on punishment.67 In A Theory of Justice, Rawls noticed that utilitarianism

66 Ibid., p. 9.

67 Ibid., p. 9n6.
dominated political philosophy almost by default, because only it was robust enough to
guide politics. Similarly, Murphy notes that utilitarianism’s dominance of punishment
theory is again almost by default, since only it seems able to provide a justification of
punishment that fits with liberal theories of the state. As we have seen, Rawlsians who
reject Nozick’s version of social contract theory might be able to justify Kant’s version of
retributivism, but since that theory is highly suspect on a number of grounds, it is not
much of a solution.

4.8 Punishment and Moral Education

There is another family of punishment doctrines that, its proponents argue, is
compatible with the values underlying a liberal state. These hold that the moral
instruction and improvement of wrongdoers is both the goal of punishment and a
legitimate goal of the liberal state. Herbert Morris calls his version the ‘paternalistic
theory of punishment,’ Nozick prefers ‘teleological retributivism,’ but Jean Hampton’s
name, “the moral education theory of punishment,” seems to be the most popular in the
literature.68 Anselm’s theory of punishment belongs to this family, as we will see in the
next chapter.

The moral education theory holds that the moral instruction and improvement of
wrongdoers is both the goal of punishment and a legitimate goal of the liberal state.

Punishment, on this view, is a complex act in which the power in authority P communicates at least the following message to the subject of punishment S:

P believes strongly that S’s act A is wrong; that S is responsible for choosing to A, knowing that it is wrong; that the rationale V underlying the law which S violated is an objectively valuable good; and that S’s decision to A makes manifest that S does not appreciate the good of V.

It is not sufficient to communicate this message verbally, for at least two reasons. The first is that if the person punished recognizes the wrong she has done and repents, she will want to do something to remove the guilt she now feels for her past conduct. Punishment provides closure and puts the guilt in the past, so that the guilty person can restore the relationships she damaged by her wrongdoing. As Morris puts it, “The punitive response to the breach defines a limit to separation that is occasioned by wrongdoing. The debt is paid, life can go on.”

Hampton develops a second reason that the message of punishment cannot be communicated verbally, but requires some sort of suffering. Because the punishable action is more serious than other prohibited actions, the authority who punishes needs a more serious way to communicate that it is forbidden. Since, Hampton assumes, the person who acts in the forbidden way is pursuing his pleasure or own ends, punishment should be painful or at least unpleasant. The ‘suffering’ in punishment need not be physical suffering, but must at least involve the ‘disruption of the freedom to pursue the

\[\text{(footnotes)}\]


70 Hampton, “Moral Education Theory”, pp. 225ff. This aspect of the moral education theory is called into question by recent studies of ‘restorative justice’ techniques of sentencing, as we will see later.
satisfaction of one’s desires.’ If it does this, the offender will take the punishment seriously, because it frustrates his most important interests (freedom and pleasure). This is like fencing off or prohibiting an action, the way an electric fence places limits on what the cattle can do without experiencing pain; it communicates at least that society forbids the action. But because human beings are more rational than cattle, there is a chance that punishment can communicate more. Ideally, it communicates that this is what you did to the other person, you hate it, and so if you consider how your victim must feel, it is reasonable to think that the other person hates it too. To someone willing to harm another, this sort of teaching by experience is the only way to communicate the gravity of the wrong.

Morris argues that something like this happens whenever a mother punishes her child. It is not enough for the mother to express disapproval of her child, or simply to make her feel guilty. In order for the child to acquire several important moral concepts, it is necessary to punish the child. For example, Morris argues, punishment vividly teaches children that there are limits on conduct, that not everything we do is right, and that doing wrong grants another the right to respond to your action in a painful way. Punishment also teaches children just how deeply parents are committed to the principles or values underlying the limit. Children naturally strike out when they or what they care

71 Cf. Morris, “Paternalistic Theory,” p. 101: “The parent’s spontaneous anger or disapproval or blame causes the child distress. They may motivate future compliant conduct. They may arouse in the child guilt. They are not, however, by themselves requital for wrongdoing and by themselves do not relieve guilt. My view is that punishment has some special and logical relationship to wrongdoing and to the possibility of a child’s acquiring the concept. Because of this relationship, punishment is connected with the good that I have described in a way that blame or disapproval by themselves are not.”
about are injured, so they can understand that when what their parents care about is injured, their parents punish. That there can be different degrees of punishment teaches children that there are different degrees of wrongness (at least in their parents’ eyes), and that there can be different degrees of fault/responsibility for breaking them (compare with the legal concept of *mens rea*). Appropriate punishments can communicate the particular character of the wrong done. As Morris notes, “Even young children will find it particularly fitting to penalize a cheater by not permitting, for a time at least, further play, for such punishment conveys the central importance of honesty in the playing of the game and one’s placing oneself outside the community of players by dishonesty.”

We’ve already noted that punishment provides closure to the wrongdoing, returning matters to where they were before and restoring relationships. We need only remark that this is especially important for children, because it is especially important that the relationship with the child be restored after the lesson has been learned. Mere assigning of guilt does not do this—if I acknowledge that I am guilty but cannot remove

---

72 How does one teach about the variety and hierarchy of wrongs without punishing? It seems impractical, for instance, to yell really loudly at the child when he has done a serious wrong while merely reprimanding her softly for a mild wrong. For one thing, it is not clear how to match the yelling with the wrong when the wrong is somewhere between the extremes—does one simply modulate decibels, or vocabulary to fit the crime? How does one calibrate this? How is the child supposed to know what lesson to take away from a thirty decibel yell versus a thirty-five decibel yell? How is the child to know that there is a rationale to the yelling, when the yelling might just be some sort of release as when mom slams her finger in the drawer?


74 Ibid., pp. 102-3: “The child’s response to wrongdoing by feeling guilt, its willingness to accept some deprivation, and its commitment to acting differently in the future, all play an indispensable role in its restoring relationships it has damaged, relationships with others and with itself.”
this guilt, by my own standards I am less of a person than I want to be, and there is no way for me to become the person I want to be.

Morris concludes from this that the complex moral message that punishment communicates is ultimately good for the child. Indeed, it is the child’s good that motivates the parents to punish, and parents who allow their children to misbehave without punishing them are usually considered lazy or negligently self-absorbed. Morris dryly points out that “It would be perverse if the parent were generally to punish primarily from motives of retributive justice or optimal utility for the family.”

Of course, it will immediately strike people as anti-liberal to apply this ‘paternalistic theory of punishment’ to adults. Liberalism does not usually describe the state as a parent or the citizen as a child. Advocates of this theory have ready answers to the obvious objections to it. And, as we have argued, it is a mistake for liberals to object to the mere idea that some adult will exercise coercive power over another unwillingly, whether or not the authority acts cruelly. For our purposes, it will suffice to note that the relationship of parent to child is an excellent instance of benign heteronomy. The parent’s goal is not domination, but the education of the child so that he or she will internalize and appreciate the values violated in committing the offence.

Morris suggests that of all the theories of punishment, this would be the most rational to choose from behind the veil of ignorance, because it is the only one which promises to treat us as fully moral and responsible persons. Murphy rejects Morris’s

75 Ibid., p. 103.
claim, for good Rawlsian reasons. However, that exchange leaves room for reconciliation between this theory of punishment and a version of liberalism amended along the lines of Anselmian mercy, which we will develop in Part Three.

---

INTRODUCTION TO PART TWO:
ANSELM’S VIEW OF JUSTICE, MERCY, AND RIGHT ORDER

In Part One, we laid out two different arguments that contemporary liberal theories need to reconsider their long-term alliance with certain views of autonomy and egalitarianism. The first argument, covering the first three chapters, made essentially two claims: 1) that a regime that tries to impose a controversial view of autonomy and egalitarianism on its citizens will be unstable, and 2) several contemporary versions of liberalism do try to impose such a controversial view. We agreed with Rawls that stability is a deeply important theme in political philosophy, and we even pushed Rawls’s argument a bit further by arguing that the promotion of stability was a political imperative.

We also argued in Part One that the same contemporary liberal theories have particularly inadequate accounts of punishment, a central function of any state. We saw that theories which try to justify punishment in terms of autonomy and egalitarianism seem rather untenable, and we began to consider a view that punishment is inherently linked to education towards certain societal moral norms. We cut short the discussion of the different moral education theories of punishment, since an extended discussion would take us too far afield. Yet even that brief discussion did bring up several points that we
will pursue in Part Three of this dissertation. The first is that theories of the state should be developed in dialogue, perhaps in reflective equilibrium, with theories of punishment. The second is that the practice of punishment requires the state to make normative moral claims, which are likely to be controversial (to some of the criminals, if no one else). A third point is that if a goal of the state in punishing is to increase stability over generations, and not just in the short-term, the practice of punishment has to be humane, with the good of the criminals being at least one of its goals. Punishment has to be heteronomous, it seems, but it also has to be benignly so in order to lead to greater stability.

In Part Three of this dissertation, we will try to show how it is possible for a state to exercise coercive dominion over its citizens while fostering in each generation of citizens so governed a principled affection for the state and its values. It might seem paradoxical to assert that a regime based on heteronomy will be more stable than one based on egalitarian autonomy, but the argument to this point has forced us to consider just this possibility. As we have suggested, a central concern of Part Two will be to develop an account of the virtue of mercy, an essential virtue for making rule by another benign.

When we consider the literature in which mercy is a theme, we can start to see how it might be possible that even a proud and independent people might develop a principled affection for those who coercively rule them. A king who shows mercy is popular with his people, while to say a ruler is merciless usually implies that he is cruel and unpopular. The Duke in Measure for Measure is viewed affectionately and favorably. Julius Caesar won great acclaim for his clementia, and clementia was considered a
signature virtue of Roman emperors thereafter. Indeed, as Seneca in his treatise De Clementia tried to convince the young emperor Nero, mercy makes rulers both honored and safe: honored because they are show self-restraint when it is within their power to be cruel, and safe, because while people will respond with love to one who loves humanity, they respond to the tyrant with rebellion. The Jewish and Christian understanding of God as forgiving and merciful is closely related to the image of God as the perfect king.

A king shows mercy to his subjects when he mitigates a sentence that has been passed, or he makes an exception to a policy that adversely affects a particular person, or he replaces the strict laws of his predecessor with ones better adjusted to the habits of his subjects. As the legal courts obtained a certain functional independence from the court of the king, mercy became associated with the law as well, so that one could throw oneself ‘on the mercy of the court.’ Theologians, who spoke of God as a king who ruled, argued that mercy would be one of his attributes. Shakespeare touches on all these traditional uses of mercy in a famous speech from The Merchant of Venice (cf. IV.1.182ff.):

The quality of mercy is not strain’d,
It droppeth as the gentle rain from heaven
Upon the place beneath: it is twice blest;
It blesses him that gives and him that takes:
’Tis mightiest in the mightiest: it becomes
The throned monarch better than his crown;
The scepter shows the force of temporal power,
The attribute to awe and majesty,

---

1 Cf. David Konstan, Pity Transformed (Gerald Duckworth & Co., 2001) p. 100: “In the year 27 B.C., the Senate honored Augustus with a shield representing the four cardinal virtues (Res Gestae 34); one of these was clemency and from this time onward, it figured prominently among the traits of emperors.”

Wherein doth sit the dread and fear of kings;  
But mercy is above this sceptred sway;  
It is enthroned in the hearts of kings,  
It is an attribute to God himself;  
And earthly power doth then show likest God’s  
When mercy seasons justice. Therefore, Jew,  
Though justice be thy plea, consider this,  
That, in the course of justice, none of us  
Should see salvation: we do pray for mercy;  
And that same prayer doth teach us all to render  
The deeds of mercy.

In this speech, Portia is trying to persuade Shylock, the Jewish moneylender, to show mercy to his debtor Antonio. Shylock insists on strict justice without mercy, exemplifying the Jew who rejects Christ and with it the picture of a merciful God. The whole speech is a classic plea for placing mercy prior to justice, appealing to the “Our Father” and St. Anselm of Canterbury’s doctrine of the atonement in the context of a civil proceeding, to show that the mercy which is so appropriate for God and king should be exercised by every man who finds that he has another at his mercy.

Shylock is unattractive because he is merciless, while the Duke is attractive because he is merciful. If we focus solely on these contrasting Shakespearean characters, we might conclude that it is always better for a ruler to be lenient than to be strict. Of course, this is far too simple, as we noted in the Introduction. Leniency always brings with it the danger of disorder, arbitrariness, and injustice. It is also possible for a ruler to

---

3 I’m attributing this view of Jewish theology to Shakespeare; it is not my own. Shakespeare’s play also raises the specter before its Christian audience of what may happen to Christians who behave with something less than ideal mercy towards their fellow human beings, as Antonio does with Shylock (cf. I.3.108ff). The contrast between (the ideals of) Jewish justice and Christian mercy makes up an important theme of Anselm’s Cur Deus Homo.
be unpopular because he is too lax, where crime runs unchecked or civil servants are lazy and self-serving.

It seems, therefore, that in order to garner the respect and affection of their citizens, the rulers of a stable society somehow have to reconcile mercy and justice, a flexible respect for the dignity and humanity of the individual person with a responsible protection of the common good. As we saw in the Introduction, St. Anselm of Canterbury developed two paradoxes of mercy and justice in his early work the *Proslogion*: a) It would seem that mercy and justice are opposed to each other. But doesn’t that make mercy unjust and justice merciless? b) It would seem that mercy is not obligatory, so that it could be shown to some people who committed a crime, but not others. But doesn’t that make mercy unfair, and thus unjust? We also saw how Jeffrie Murphy argued, quite plausibly, that these paradoxes seem to afflict very practical questions of justice and mercy even to this day, in our debates about sentencing guidelines and prosecutorial discretion. If a stable society must be characterized by both mercy and justice, solving these conceptual paradoxes becomes an important practical political matter. This is our task in Part Two.

St. Anselm not only proposed these paradoxes, but he spent decades trying to solve them. In chapter seven, we will sketch his solution in some detail. The solution comes out of Anselm’s view of punishment, and so to some extent the subject of chapter seven is the same as that of chapter four. Anselm’s view of punishment assumes that the goal of a ruler in punishing is the protection of order, and that justice is concerned with the preservation of order. But an exclusive concern with preserving order can lead one, in the extreme, to reestablish order by punishing everyone! This, Anselm argues, is the
dilemma of the Christian God when faced with a human race marred by original sin. Justice would require that everyone be punished, but it would seem pointless for God to have created the human race simply to punish it. And yet, God cannot deviate from justice. So if God is to be merciful and refrain from punishing at least some people, His mercy must be consonant with justice and right order. Based on Anselm’s arguments, we will suggest that for him, mercy is the restoration of right order for the sake of that right order itself.

Anselm thinks that the definitions of divine attributes, e.g. Justice, are also the definitions of the corresponding human attributes. In chapter five, we will look at the philosophical tools Anselm uses to make this argument. Central to his view is the idea of a pure perfection, an attribute that it is always better to have than not to have, which excludes from its definition any reference to limitations or finite modes of being. Pure perfections can be predicated of God because they do not include any connotations of limitation or finitude. They can also be applied to creatures, and so a pure perfection can be defined so as to be common to both creature and Creator. Anselm’s solution to the classic problem of the divine names is importantly different from the theory of analogy developed by Aquinas, in part because Anselm develops the method of concept purification in the service of a Neoplatonic ascent to God rather than a scholastic separation of the sciences.

There is very little published on Anselm’s theory of concept purification, and almost nothing on his speculative views regarding punishment and governance. In chapter six we will address another important lacuna in the Anselm scholarship. Anselm is not often considered to be a theorist of justice, or to have had much to say about the
topic. The burden of chapter six is to show the progression of Anselm’s thought on these matters, from his very early *oratio* to St. Paul, to his first speculative work the *Monologion*, through the *Proslogion*, his three treatises on truth, free will, and the fall of the devil, until his masterpiece the *Cur Deus Homo*. We will show a consistent concern with justice, a development of the paradoxes of justice and mercy, and the assembly of the solution to those paradoxes in his development of the ideas of rectitude and freedom of the will, all of which get brought together in the *Cur Deus Homo*. Chapters five and six might seem like an interlude from our main argument about liberal accounts of justice, but they are necessary to help us understand Anselm’s solution to the paradoxes of justice and mercy, and the account of liberalism we develop in Part Three depends on Anselm’s view of justice developed in response to these paradoxes.

In the Part Three, we will return to our discussion of contemporary liberalism and the role of mercy in it. We can already suggest some features of a liberalism based on mercy, and it will be helpful to keep these in mind as we develop our arguments about Anselm’s political views. Like any liberalism of fear, a liberalism based on mercy would hold that the prevention of cruelty is a fundamental value, and so it will require institutions and practices to prevent all cruelty, whether by the state or by non-governmental institutions. Powerful institutions, whether state-run or not, should be subject to checks and balances of some sort, including criminal penalties for abuse of power. This is not to spread an atmosphere of mistrust, but quite the contrary: the knowledge that there are well-functioning mechanisms to hold persons accountable for
how they act in positions of power allows all citizens to trust that the powerful persons will behave themselves.\(^4\)

This liberalism would not strive for the more ambitious goal of protecting autonomy, but would settle for a state that is what we called benignly heteronomous. Therefore this liberalism would not need to require that all public institutions and actions be justifiable to all citizens, but only that they respect the dignity of those people affected by the institutions and actions. Public decisions do not need to be transparent, but they should be open where possible, and when openness is not possible they should be subject to democratic oversight and accountable to the public in some fashion.

In abandoning the goal of autonomy, though, this liberalism is also forced to abandon tolerance as it has been articulated in liberal thought since Rawls. Tolerance is what we might call a horizontal virtue—it is appropriate among equals, but is not appropriate in cases of inequality. Rawls is the clearest about this feature of tolerance: citizens in a well-ordered society regard themselves as equal to each other, and justice is to be considered as fairness, given this equality.\(^5\) But in cases where one person or group is given political power over another, it might not be the case that the lower person or group (the ruled) ought to be given the same liberties and powers as the higher (the rulers). The Supreme Court, for instance, has much more latitude for interpretation than do the lower courts, which must toe the line set by the highest court. A child is not given the same freedoms as a parent (even legally). The CEO has more freedoms (e.g., he can

\(^4\) Rawls makes this point effectively in TJ, §§38-9.

\(^5\) Cf. PL pp. 18-9; TJ §§3, 33.
take a longer lunch, or start work when he wants, without fear of being fired) than his employees. This is not to say that the ruler might not choose to rule in an egalitarian fashion (there was a laissez-faire movement in parenting some years back, e.g.), leaving open the possibility that, at one extreme, this liberalism might look very much like Rawlsian political liberalism in certain circumstances. However, in cases where there is an inequality in power or in responsibility, there is a need for a vertical virtue that will ensure that these conditions of heteronomy remain benign. I argue that Anselmian mercy is this virtue.

Michel de Montaigne, whom Shklar names as one of the founders of the liberalism of fear, agreed with Portia’s view that everyone, and not just the king, should show mercy. His Essays were written as a rebuttal to Machiavelli and those who took his advice that cruelty was permissible if political necessities so dictated. Montaigne’s mercy tends to be less of a virtue than a skeptical suspension of judgment whenever someone argues that cruelty is necessary to advance some allegedly just cause. Montaigne tries to show that, in his case, skepticism has made him aware of the feebleness of his own powers, and has led him to a mild manner in dealing with the weaknesses of others; thus he commends skepticism to others. Obviously, a mercy grounded in skepticism cannot be acceptable to integralists, and so cannot be the virtue we are seeking.

6 This is the argument of David Quint in Montaigne and the Quality of Mercy: Ethical and Political Themes in the Essais (Princeton University Press, 1998). It should be noted that the Essays, which are often thought to have invented a literary genre, do show a striking stylistic as well as topical resemblance to the Discourses on Livy.
Anselm’s view of mercy, on the other hand, is not based on skepticism or on mildness of disposition. This is because Anselm sees mercy first of all as an attribute of God, who actually does know all, and who does not have dispositions, mild or otherwise. Anselm’s view of mercy has some similarities with the Stoic idea of *clementia*, in that it depends on right judgment rather than disposition or passion. But it far surpasses the Stoic idea in its depth and power, and in its claim to being normative rather than merely a good tactic for an emperor to win favor with his subjects. Anselm’s idea of the mercy of God is deeply influential wherever Latin Christianity spread, and as the quote from *The Merchant of Venice* suggests, its influence extends beyond theology to law and politics.

Another part of the task of developing a mercy-based liberalism will be to fill the hole in liberal theory left by discarding the idea of autonomy. Two worries about benign heteronomy immediately suggest themselves: first, the ruling groups or persons are likely to become paternalistic; second, the ruled are likely to become irresponsible and complacent. These tendencies of heteronomy are correlative, and reinforce each other. The ruler notices that he has the ability to provide some service to the ruled, and does. The ruled thus are not forced to acquire the desired service, and so either lose or never develop the ability to do so, thus creating a need for the ruler to act more often. The idea of autonomy, on the other hand, gave the individual not just freedom but also personal responsibility, or at least, it absolved anyone else of responsibility for another’s actions.

---

Any stable regime will need to retain at least this much of the ideal of autonomy, the sense of personal responsibility.

Anselm’s account of mercy does this. For Anselm, mercy only makes sense when combined with an account of responsibility. For him, justice is reducible to ‘rightness,’ in the sense of the right ordering of that for which one is responsible. In my sphere of responsibility, justice is making sure that everything is ordered correctly. This ordering might extend beyond externals to Locke’s “care of souls”—some people (e.g., abbots, parents, bishops) might find themselves with the duty of caring for the souls of others, in which case they have to find a way to order the souls of those for whom they are responsible. Anselm believes that all people have free wills, and that even God cannot override this freedom. So just as with the ideal of autonomy, every person has responsibility for the right ordering of his own soul, at least to some extent. Unlike with autonomy, however, that responsibility might be limited—in some cases it might just be to obey one’s superior with docility. The superior in that case has to make decisions that will be for the others’ own good, trusting in the others’ willingness to follow the decision freely.  

Where the superior has responsibility for the souls of the others, he has to train those souls in virtue. He must do what he can to persuade each soul to love as it ought and to will as it ought. It is not sufficient to get everyone to behave appropriately, for

---

8 This is the division of responsibility in the Rule of St. Benedict. The abbot is responsible for the salvation of the souls of his monks, and the monks are responsible for obeying the abbot. If the monks disobey and resist correction, the abbot is absolved of responsibility for their souls. Otherwise, he will have to answer to God for their failings. See the discussion of the political view of St. Benedict below in chapter seven.
people can behave out of fear or a selfish desire to get ahead. They must behave for the
right reasons. Therefore, it would be a mistake for the superior to slide towards
autocracy, for that would do a long-term disservice to the souls under his care. Anselm’s
theory of responsibility, combined with his theory of freedom and the will, requires that
the superior never rule out of fear, but instead try to win over those subject to him so that
they willingly follow his leadership. As we shall see, when the parts of Anselm’s theory
work together, they make for a surprisingly modern view.

So in the next chapters, we will look very closely at Anselm’s writings, tracing
the themes of justice, mercy, freedom, responsibility, and the will as they develop
throughout his corpus. After that is done, we will be able to develop our Anselmian
liberalism, and draw out some of its implications.
5.1  Anselm on God’s Mercy and Justice: An Overview

Traditional Christian theological language lists both justice and mercy as attributes of God. These two attributes are usually paired, for obvious reasons. If we pay too much attention to the justice of God without paying attention to his mercy, then we end up with a picture of a wrathful God making strenuous demands on people and punishing them when they (almost inevitably) fail. But if we neglect God’s justice in favor of his mercy, then we end up with an enabling God, a pushover who, because he is always and infinitely forgiving, fails to distinguish among the good and bad, to treat similar cases similarly, and to uphold high standards of behavior. One picture makes God merciless, while the other makes him unjust.

But if we say that God is both just and merciful, then we have another problem. For God is not simply just, but Justice itself; nor is he simply merciful, but Mercy itself. This is because Anselm subscribes to the classical doctrine of divine simplicity, which holds that all the attributes of God must not be really distinguishable from the divine essence. For any attribute of God, God not only has it, He is it essentially. This strong requirement obviously prohibits attributing contrary properties of God. If mercy and justice are defined in opposition to each other (e.g., justice is strictly enforcing the law
while mercy is being more lenient than the letter of the law prescribes), then God cannot be both Mercy and Justice. If justice and mercy are opposed, then for God to be Justice He cannot be Mercy, and vice versa.

Anselm’s solution is to reconceive justice and mercy so that they are expressed in terms of a more fundamental idea. Divine justice and divine mercy are two different ways to understand the providential order God has willed for the universe. So justice is the name for the state of affairs where the universe is ordered according to God’s plan. Mercy foregrounds the idea that in his plan God is attentive to creaturely limitations and possibilities, respects the freedom of all rational creatures, and works to help all creatures fulfill their natural ends. Because God is omniscient and omnipotent and eternal, he has providentially ordered every aspect of creation, even to the point of accounting for certain creatures’ sinful abuse of their freedom.

This means that God’s plan for the cosmos is not simply a static vision of a perfect state of affairs—indeed, his providential ordering is quite dynamic. God does have a perfect state of affairs in mind as his final goal, but that is no more than the proper state of affairs at what we might call time $T_\infty$. The proper state of affairs at $T_{1971\text{AD}}$ is different from that at $T_\infty$, but is a stage in the movement of all creation towards that state. Justice and mercy both demand that at this time the world be moving towards its ultimate state of affairs. This ‘moving towards’ is not Whiggish or asymptotic—Anselm did not suscribe to the view that human imperfections would disappear with the progress of time. But he did think that there was a difference between the state of affairs that is fitting for the present circumstances and that which is fitting simpliciter. For people to want the world to be perfect right now would be against God’s will for this moment; it is more
appropriate to want the world to be as it ought to be now, and perhaps to be a little better tomorrow. In Anselm’s view, mercy emphasizes where things are, while justice points towards the vision of what must come to be. To say that God is merciful and understanding emphasizes that he is not asking more than weak creatures can give at this stage. To say he is just and demanding is to remember that he expects creatures to give all they can.

This way of talking about justice and mercy as attributes of a perfect God was not the only, or even the dominant framework in Anselm’s day. Anselm lived during the era of the “Gregorian revolution” (after Pope Gregory VII of Canossa fame), a time when the Church started to distinguish itself institutionally and juridically from the rest of society. Gregory and the other eleventh century reforming popes used the law as a principal means of their institutional and societal reform, setting up an extensive court system all over Europe and sponsoring intense studies of Roman and canon law.¹ In this climate where the law was very prestigious (a few decades after Anselm’s death, a well-regarded lawyer might be the object of bidding wars between cities, just like prize athletes are today!), justice was often construed in legalistic terms.² Feudalism provided another set of ideas about justice and mercy. Lords had their vassals “at their mercy,” unless the vassal could appeal to some “right” (jus) to act in a certain way, which right was derived either from a contract with the lord, from the feud between lord and vassal, or by appeal


to some broader principle of right.\(^3\) There was a popular idea of the Atonement before Anselm that demonstrates this idea: After the fall, God gave the devil the right to the souls of all sinful men (alternatively, in the fall, man voluntarily made himself the bondservant himself to the devil). God had to respect this right, being just, but because of his mercy he still wanted to save men’s souls. So God became man in order to trick the devil. The devil took the bait when he claimed the soul of the innocent Christ, thereby overstepping his prerogatives and losing his right to the souls of those who attached themselves to Christ.\(^4\)

Although Anselm was not above using feudal metaphors to express his ideas, he was too much the philosophical theologian to rely on such metaphors to guide his thought. If the devil stole man from God, God owes him nothing. If man swore his allegiance to the devil, it was an allegiance that he did not have to give, since he owed it first to God. Most fundamentally, if God is omnipotent, the devil has no power over Him, and so God need not resort to trickery to attain what He desires. From his very first systematic treatise, the *Monologion*, Anselm treats justice as an attribute of God first, rather than as a feature of law. Later, under polemical pressure from Jewish and Muslim critics of Christianity, he seems to have used the tools he developed in natural theology to show just how far this attribute of God is compatible with the understandings of mercy and especially justice common in his day, and where the contemporary understandings

---


needed to be revised in light of natural theology. In this chapter, we will examine these tools, and in the next chapter, we will see how Anselm deploys them in his discussion of justice and mercy.

5.2 Tools of Natural Theology: Anselm on the Attributes of a Simple God

Isidore of Seville (d. 636) includes the following discussion in his *Etymologies*:

The word ‘merciful’ is derived from the suffering along with *(conpatiendo)* the fate of another’s misery; and from this ‘misericordia’ is named, since the misery of the suffering person makes the merciful person’s heart suffer *(miserum cor)*. Yet this etymology is not found everywhere; for mercy is in God without the suffering of his heart.

If mercy means to have a suffering heart, as its Latin name suggests, then it would seem inappropriate to say that God is merciful, since God does not have a heart and traditionally is not thought to undergo any change, let alone suffering. But a Christian is committed to believing (quite deeply) in God’s mercy. Anselm expresses the problem in similar terms:

5 Isidore of Seville, *Etymologiae*, 10.164: Misericors a conpatiendo alienae miseriae vocabulum est sortitus: et hinc appellata misericordia, quod miserum cor faciat dolentis aliena miseria. Non autem occurrerit ubique haec etymologia; nam est in Deo misericordia sine ulla cordis miseria. (My somewhat loose translation.)

6 For a discussion of this tradition and a defense of its philosophical and theological power, see Thomas G. Weinandy, *Does God Suffer?* University of Notre Dame Press, 2000.

7 P.8. (1.106): Sed et misericors simul et impassibilis quomodo es? Nam si es impassibilis, non compateris; si non compateris, non est tibi miserum cor ex compassione miseri, quod est esse misericordem. At si non es misericors, unde miseris est tanta consolatio?
But how are you [God] simultaneously both merciful (misericors) and impassible? For if you are impassible, then you are not suffering along with another (conpateris); but if you are not suffering along with another, then yours is not a heart made to suffer by suffering along with another’s misery (miserum cor ex compassione miseri), which is what it is to be merciful (misericordem). And if you are not merciful, whence comes so much consolation for the sorrowful?

This is the familiar problem of how to characterize the attributes of a simple God, what some call (following Pseudo-Dionysius) the problem of divine names.\(^8\)

The problem of divine names faces anyone who holds (1) that signification is by means of concepts, (2) that concepts are likenesses of the things conceived and are more or less true insofar as they are an essential likeness of the thing, and (3) that no created intellect can form even a remotely approximate likeness of God’s essence. Anselm joined Boethius and Augustine, who also faced this dilemma, in deciding that the best way to improve our conception of God is to understand His attributes. These attributes can act as a sort of bridge between our understanding of the world around us and God, since if we can form concepts by which we can understand some of God’s attributes, we might be able to assemble them so that taken together as one, they formulate a complex concept of God that more closely approximates his nature. The problem of understanding God’s essence is thus exchanged for the apparently more manageable problem of understanding the divine attributes.

But as our discussion of God’s misericordia shows, the divine attributes are not much easier to understand. One set of problems arises when the attributes have

connotations that are incompatible with God’s immateriality; as we have seen, since God does not have a heart (cors), it is not clear how he can have the attribute of misericors.

But a deeper set of problems arises whenever any attribute of men is applied to God, even if (as in the case of justice, goodness, and love, inter alia) the attribute makes no essential reference to qualities of physical bodies. To appreciate this problem in more detail, however, we will need to look more closely at Anselm’s account of concept formation and signification.

5.3 Concepts as Mental Words

In chapter 10 of his earliest work, the Monologion, Anselm distinguishes two types of verba (“words”): one kind of verbum signifies the thing conventionally, the other signifies naturally. Conventional signification occurs when I say (or even think to myself) the word “man” and signify man; in this case, the written or spoken or silently pronounced word is associated with the thing it signifies by convention, explaining why different languages use different names to signify the same thing. The other sort of verbum Anselm calls a natural verbum, since it signifies what it signifies naturally, that is, for all men and in all languages.

By ‘natural verbum’ Anselm actually means a concept, which is called a verbum only by analogy. Such a verbum is a mental “word”, a conception of the mind (conceptio mentis) that is “spoken” mentally.⁹

⁹Monologion (M), ch.10. My translation. “Mentis autem sive rationis locutionem hic intelligo, non cum voces rerum significativae cogitantur, sed cum res ipsae...acie cogitationis in mente conspiciuntur.” Anselm’s mixed metaphor is quite common in the Christian tradition, where the second person of the Trinity is called both the Word (Jn.
Here I understand by “rational” or “mental” speech, not when the words significative of things are thought, but when the things themselves are viewed in the mind with keenness of thought…

Anselm thinks that mental concepts are images or likenesses of things, like the image of the product that the craftsman has in his mind before he makes it. He says that we form these concepts in two ways: by “imagining” (forming a mental image of) sensible things, or, whether something is sensible or not, by understanding its formal account (ratio) or universal essence (universalis essentia), i.e., what is expressed by the definition.\(^\text{10}\)

Anselm, following Aristotle, declares that mental words “are natural and are the same for all peoples.”\(^\text{11}\) The reason that these verba are the same in all languages is that a verbum is a mental conception or image, which is formed by those processes of cognition that are common to all men; spoken or written words, on the other hand, are related to these natural verba only by convention. Because a natural verbum is an image or likeness in the mind of some thing outside the mind, such a verbum can be said to be more true the

\(^{1.1}\) and the Image (e.g., Heb. 1:2-3; Col. 1:15-17) of the Father. Augustine, for instance, describes “a word that is before all sound, before all thought of a sound” as being “the most like to the thing known, from which also its image is begotten...when it is a word belonging to no tongue, but is a true word concerning a true thing, having nothing of its own, but wholly derived from that knowledge from which it is born” (De Trinitate, 15.12.22). Boethius and Anselm, following not only Augustine but also Aristotle’s declaration that “vocal sounds are symbols of impressions in the soul...[which] are the same for all” (De Interpretatione, ch.1, 16a4-9), determined that it is fitting that what is within our soul and is signified by our spoken and written words be called a ‘verbum’.

\(^\text{10}\) Ibid.

\(^\text{11}\) M, ch. 10; translated by Jasper Hopkins and Herbert Richardson in vol. 1 of Anselm of Canterbury. (New York: Edwin Mellen Press, 1976) p. 19: “...naturalia sunt, et apud omnes gentes sunt eadem.” Where indicated by (H&R), the translations are taken from these three volumes of translations. Where there is not indication, the translations are mine.
more closely it corresponds to the object of which it is the likeness. It follows then that the more one learns about the object so conceived, and the better the mental image formed, the more true the word is.

What Anselm means by understanding (intellectus) is the act of forming a better natural verbum. It is impossible for any creature to understand God fully in this sense, because it would require that a creature form a verbum that completely expresses God’s ratio or universal essence, something that only the second person of the Trinity, the Verbum Dei, can do. It is very difficult for us to form any mental verbum of God at all; not only can we never sense Him directly, but he is so much greater than we are that no concept that can ever exist in our minds could ever be a very good likeness of His nature. As Anselm declares in his Proslogion, God is greater than we are able to understand.

5.4 Mental Words and the Limits of Understanding

Anselm’s natural theology proceeds by taking the verba that apply to things we do understand and which can also be applied to God, and leveraging what understanding we do have in order to increase our understanding of God to whatever small degree.

Ibid.

Cf. Epistola de Incarnatione Verbi (DIV), ch.1: “...he who does not believe, does not understand. For whomever does not believe, does not experience, and he who has not experienced does not learn from experience. For insofar as experience is superior to just hearing of a thing, so knowledge from experience surpasses knowledge from hearing.” “...qui non crediderit, non intelliget. Nam qui non crediderit, non experietur; et qui expertus non fuerit, non cognoscet. Quantum enim rei auditem superat experientia, tantum vincit auditentis cognitionem experientis scientia.”

Prosplogion (P), ch. 15.
Following Aristotle’s De Interpretatione, Anselm holds that a spoken word signifies what it does by means of the *verbum*. As said above, one of the ways that one forms a natural *verbum* of something is by understanding its formal account (*ratio*) or its essence, so to whatever degree such a *verbum* approximates its object, to that degree the essence will be signified by means of the *verbum*. Therefore, the natural *verba* place limits on signification; if someone were to signify Venus by ‘the morning star’, he signifies Venus only insofar as it is the morning star, and does not signify, e.g., that it is the evening star. Were he to have signified Venus by ‘Venus’, if his concept of Venus included that it appears in both the morning and the evening, both features would be signified, although only indistinctly. Anselm describes this for the case of ‘man’.\(^\text{15}\)

Indeed the name ‘man’ signifies *per se* and as one thing those things from which the whole man consists. Among these things substance holds the principal place...

So when a term signifies *per se*, it signifies the essence chiefly, but it also signifies the entire object. In the case of ‘man’, for instance, the essential definition would be *rational mortal animal*, while the word might also imply certain other characteristics of man that are accidents but that are always present, such as being able to laugh, having weight and height, being non-translucent, being made in the image of God. These are not included in the definition, since definitions only capture the essence, but they are included in the signification of the word. They are included, however, not as individuated characteristics, but “as one” (*ut unum*). This is because in signifying *per se*, a written or spoken word

\(^{15}\text{De Grammatico (DG), ch. 12 (my translation): Nempe nomen hominis per se et ut unum significat ea ex quibus constat totus homo. In quibus substantia principalem locum tenet...}
appeals to the *verbum* in order to signify ultimately its object. The *verbum*, if it corresponds to its object closely enough, might then include certain characteristics that are not captured by the definition, but that the mind associates with the object nonetheless.

According to Anselm, when believers use words that signify God, they actually signify Him only to the degree that their concepts are like Him.\(^{16}\) If my concept of God is of an old man with a white beard, when I say His name I signify Him insofar as he is an old man with a white beard (which might mean, if God is completely unlike my concept, that I don’t signify *Him* at all). Since we cannot perceive God directly by our senses, if we can form a *verbum* of Him, it will be because of revelation or else because of some argument from what we do know about the sensible world to what God must be like as a result. Anselm’s project, which rules out the former option, means his theological language must be based in concepts drawn from non-theological human experience.

This creates a problem. An inadequate concept of God will mean that any argument or syllogism that uses ‘God’ as one of its terms is likely to be invalid, or at least its conclusions will be about something that is not exactly God. But since, as we have seen above, no human *verbum* can be more than a dim likeness of God’s nature, it seems that all concepts will be inadequate concepts, in this sense. Anselm’s project appears destined to run into insurmountable difficulties of this sort.\(^{17}\)

\(^{16}\) Cf. M 65

\(^{17}\) In M 65, Anselm alludes to 1 Corinthians 13:12 in admitting that his arguments may allow him to see God only ‘through a glass darkly.’
5.5 Anselm’s Purification Procedure

Gaunilo advances a version of this objection in *Pro Insipiente*. He argues that if someone describes to him a man he had never known, he could form a concept of him by reference to what he already knew about men.\(^\text{18}\)

Now, suppose that I were to hear something being said about a man totally a stranger to me—[a man] whom I was not even sure existed. Still, by means of the specific or generic knowledge by which I know what a man is (or what men are), I would be able to think of him as well, by reference to the very thing that a man is. However, it could happen that the one who told [me about this stranger] was lying and that the man whom I thought of does not exist. Nonetheless, I would still have thought of him by reference to the true [i.e., real] thing which any man is (though not which that man is).

Gaunilo claims that on Anselm’s view, it is not possible to form a similarly vivid picture of God. It is not possible to know God directly in this life, and God, who is not in a genus or species, is so unlike anything else, by Anselm’s own admission, that it would be impossible to make inferences about Him from similar things better known to us. Anselm’s description of God as ‘that than which nothing greater can be conceived’ merely leads the fool to assemble a vague concept pieced together from the images taken from the words in the description, which taken as a whole provide little or no real knowledge of the object to which the word ‘God’ allegedly refers.\(^\text{19}\)

\(^\text{18}\) *Pro Insipiente*, ch. 4 (H&R): Nam si de homine aliquo mihi prorsus ignoto, quem etiam esse nescirem, dici tamen aliquid audirem: per illam specialem generalemve notitiam qua quid sit homo vel homines novi, de illo quoque secundum rem ipsum quae est homo cogitare possem. Et tamen fieri posset, ut mentiente illo qui diceret, ipse quem cogitarem homo non esset; cum tamen ego de illo secundum veram nihilominus rem, non quae esset ille homo, sed quae est homo quilibet, cogitarem.

\(^\text{19}\) Ibid.
Anselm responds to this in two ways, which I will treat in reverse order. In chapter 9 of his response to Gaunilo, Anselm points out that he does not require that the fool comprehend God’s nature, but merely that he apprehend what the description means and implies:

Yet, even if it were true that that than which a greater cannot be thought could not be thought or understood, nonetheless it would not be false that [the phrase] “that than which a greater cannot be thought” can be thought and understood. Nothing prevents our saying [the word] “ineffable,” even though that which is called ineffable cannot be said. Moreover, we can think [the concept] unthinkable, even though that which merits to be called unthinkable cannot be thought. By the same token, when “that than which nothing greater can be thought” is uttered, without doubt what is heard can be thought and understood, even if that thing than which a greater cannot be thought could not be thought or understood.

Whether or not this helps Anselm save the ontological argument, it is clear that his epistemological claim is more modest than Gaunilo realizes. He is not trying to grasp the nature of God with his description; he is simply trying to get some (not entirely negative) description to stick. Anselm agrees with St. Paul that in this life the believer can know God as “through a glass darkly”—that’s not much, but it is more than nothing. He also cites St. Paul in Romans 1:20 to establish that “the invisible things of God (including His

\[\text{\textsuperscript{20}}\text{Response to Gaunilo, ch. 9 (I’ve made a couple stylistic changes to the H&R translation): Sed etsi verum esset non posse cogitari vel intelligi illud quo maius nequit cogitari, non tamen falsum esset “quo maius cogitari nequii” cognitari posse et intelligi. Sicut enim nil prohibet dici “ineffabile”, licet illud dici non possit quod “ineffabile” dicitur; et quemadmodum cogitari potest “non cogitabile”, quamvis illud cogitari non possit cui convenit “non cogitabile” dici: ita cum dicitur “quo nil maius valet cogitari”, procul dubio quod auditur cogitari et intelligi potest, etiam si res illa cogitari non valeat aut intelligi, qua maius cogitari nequit.}\]

\[\text{\textsuperscript{21}}\text{M 65, making reference to 1 Corinthians 13:12.}\]
eternal power and divinity) are clearly seen from the creation of the world, being understood through those things that have been made.\textsuperscript{22} He thinks this passage indicates that he should be able to apprehend something of God by inference from the things of this world, despite the difficulties. And he thinks he has a method for doing just that.

In chapter 8, Anselm argues against Gaunilo’s point that God’s transcendence makes him so alien to our experience that he is beyond comparison.\textsuperscript{23}

Moreover, you maintain that upon hearing of that than which a greater cannot be thought you cannot think it (or have it in the understanding) by reference to any object known to you through species or genus. For [you claim that] you are neither acquainted with this thing itself nor able to make inferences about it on the basis of some other similar thing. Yet, the facts are clearly otherwise. For every lesser good, insofar as it is a good, is similar to a greater good. Therefore, to any rational mind it is clear that by ascending from lesser goods to greater goods, we can—on the basis of those things than which something greater can be thought—make many inferences about that than which nothing greater can be thought.

The “insofar as” here is important—Anselm’s argument seems to be that for any A and B, in whatever sense B is greater than A or A is less than B, A and B are similar enough for comparison. That than which nothing greater can be conceived must then have some similarity with everything else that can be compared to it. Since Anselm is not trying to comprehend God, but merely apprehend Him, even a tenuous connection between human experience and the Divine nature might help him think and speak cogently about God.

\textsuperscript{22} Response to Gaunilo, ch. 8.

\textsuperscript{23} Ibid. (H&R): Item quod dicis ‘quo maius cogitari nequit’ secundum rem vel ex genere tibi vel ex specie notam te cogitare auditum vel in intellectu habere non posse, quoniam nec ipsam rem nosti, nec eam ex alia simili potes conicere: palam est rem aliter sese habere. Quoniam namque omne minus bonum in tantum est simile maiori bono inquantum est bonum: patet cuilibet rationabili menti, quia de bonis minoribus ad maiora conscedendo ex is quibus aliquid maius cogitari potest, multum possumus conicere illud quo nihil potest maius cogitari.
Anselm thinks he can dismiss Gaunilo’s apophatic worry that there can be no basis for comparison between God and sensible creatures. Indeed, he develops a method for using those similarities between God and creatures to modify verba taken from created things so that they can be fittingly applied to God, what we will call his *purification procedure*:  

Who, for example, cannot think of this—even if he does not believe that what he thinks of actually exists—namely, that if something which has a beginning and an end is a good, then a good which although it begins does not cease is much better. And just as this latter is better than former, so also that which has neither a beginning nor an end is better than this even if [the third] is always moving from the past through the present toward the future. Yet, (whether or not there exists in reality some such thing) that which does not lack anything at all, nor is compelled to be changed or moved is far better still than this [third good, which does change]. Can this last [unchanging good] not be thought? Can anything greater than it be thought? Is not this [procedure the same as] making inferences—on the basis of those things than which a greater can be thought—about that than which a greater cannot be thought? Therefore, there is a way to make inferences about that than which a greater cannot be thought. In this way, then, the Fool, who does not accept sacred authority [i.e., Scripture], can easily be refuted if he denies that on the basis of other things inferences can be made about that than which a greater cannot be thought.

The procedure is roughly this: Anselm lists an attribute of ordinary good things, viz., having a beginning and end, recognizes that this attribute is a limitation on the good

---

24 Ibid. (again, I’ve tweaked the translation of H&R): Quis enim verbi gratia vel hoc cogitare non potest, etiam si non credat in re esse quod cogitat, scilicet si bonum est aliquid quod initium et finem habet, multo melius esse bonum, quod licet incipiat non tamen desinit; et sicut istud illo melius est, ita isto esse melius illud quod nec finem habet nec initium, etiam si semper de praeterito per praesens transeat ad futurum; et sive sit in re aliquid huiusmodi sive non sit, valde tamen eo melius esse id quod nullo modo indiget vel cogitur mutari vel moveri? An hoc cogitari non potest, aut aliquid hoc maius cogitari potest? Aut non est hoc ex iis quibus maius cogitari valet, conicere id quo maius cogitari nequit? Est igitur unde possit conici “quo maius cogitari nequeat”. Sic itaque facile refelli potest insipiens qui sacram auctoritatem non recipit, si negat “quo maius cogitari non valet” ex aliis rebus conici posse.
thing, and so imagines another being that does not have the limitation. He claims that the less-limited being will always be better than the more limited one. This may require a bit of unpacking, so let’s give a simple example.

A particularly good apple is a good thing to eat; if we were able to eat the apple and yet it were to continue in existence, then that would be a better apple than one that would cease to exist once we ate it. We don’t need to believe that such an apple exists; we need merely acknowledge that the apple would be better were it to last evermore. The same logic applies if we remove the apple’s other temporal limitation, and imagine that the apple had no beginning but existed always. Now imagine that our everlasting apple, once eaten, never ceased to exist, but that it did take time to grow back to the height of its perfection each time it was eaten. Imagine if this limitation were taken away, so that the apple always remains at the height of its perfection, so that it would always be ready to eat. Such an apple would be better than the very good apple we started with, because it would have fewer limitations. We don’t need to believe that such an apple exists, nor must we have any idea about how such an apple could exist. The point of the purification procedure is that we can still make some claims about the eternal apple based on what we know about the original apple.

Since this example only involves apples, it only illustrates the validity of Anselm’s procedure within the same species. Anselm intends that his procedure apply more generally, so that if L is a limitation on some good thing,

for any L, every good thing which lacks L (is not limited by L) is better in this respect than any good thing limited by L.

Even our eternal apple has some limitations—it cannot think, for instance—and so any human being is better than that apple, at least in this respect. Indeed, this is Anselm’s response to Gaunilo’s famous objection of the perfect island: while one might try to remove all the imperfections from an island using this procedure until one arrives at a perfect island, there are some limitations to an island qua island that cannot be removed,
providing the disanalogy with the case of God. From his description of God, Anselm concludes that God must be such that if L is some conceivable limitation on some good thing,

for every L, God is not limited by L

Anselm’s purification procedure allows him to make conceptual linkages between an unlimited being and a limited being by thinking of the former as the latter without its various limitations. If his procedure works, then he can make claims about God in human language.

5.6 Pure Perfections

As found in his Reply to Gaunilo, Anselm’s purification procedure needs some refinement. For it is clear that he cannot remove certain limitations from some good things without lapsing into incoherency. Some limitations follow so closely upon a thing being what it is that one cannot coherently think of the thing without them. Removing the limitation having a body from an apple, for instance, does not help us to think coherently about God—it doesn’t even help us think about apples, because an apple is the sort of thing that must have a body. A bodiless apple is as unthinkable as a round square; neither is a candidate for an attribute of God. The purification procedure cannot start just anywhere. It can only generate a divine attribute if it begins with the right sort of creaturely attribute, as Anselm describes in Monologion 15.

Following the precedents of Augustine and Boethius, Anselm examines Aristotle’s categories to see what sorts of attributes can be applied to God. For this

---

25 Reply to Gaunilo ch. 3.

26 Cf. De Trinitate, Book V.

27 Cf. De Trinitate, ch. iv-v.
purpose, Anselm divides the categories into three groups: substance, relation, and non-relational properties.

We have already seen why God cannot be said to be an unlimited version of some created substance, namely, that created substances are limited necessarily and essentially, and thus in their definition, so that to think of them without their essential limitations is to think a contradiction. In *Monologion* 15, Anselm makes this point regarding ‘gold’: it is not always better for a thing to be gold than not to be gold; for instance, it is better that a human being be not-gold than be gold (as King Midas learned), because it is better to be a human being, and human beings cannot be made of gold.

Relational accidents present a different problem. When we say of God that He is greater than something else, we do not signify his substance by the relational term ‘greater than’. If we did, then it would mean that there would be something necessary about God’s relation to creation, that God’s nature in some way depends on the existence of creation. But this would contradict God’s independence. Moreover, it would imply that were creation not to exist, God would be diminished in his essential power or goodness.

Anselm thinks such relational terms are important; indeed, they are the key to the *Proslogion* argument. But because they imply some limitation on God, they cannot be considered to name what God is essentially. (This fact is important for Anselm’s

---

28 M 15.

29 Anselm does not hold that it is a contradiction to call God “a substance”, but that’s because ‘substance’ names a non-relational property that it is always better to be than not to be. Likewise the word ‘God’, which is a substance-term, does not imply any limitation. The point here is that Anselm thinks it wrong to attempt to purify the names of particular substances to turn them into divine names. Cf. M 26.

30 Of course, Anselm does not think that to introduce the Trinitarian relations attributes imperfections to God. See M 29 ff.
ultimate doctrine of mercy—if mercy implies a relation to fallen creatures, then God cannot be essentially merciful.)

The third sort of term describes non-relational properties. Those terms that are predicated non-relationally are of two types. Some are such that it is always better to be described by them than not—just as it is always and in every respect better to be true than not true, to be just than unjust, to be good than not good. The other terms signify things which in some respect it is not desirable to be, e.g., sometimes it is better not to be six feet tall (if you want to fit into a small place, say) and sometimes it is better not to be temperate (since temperance is a quality whereby humans control their passions, and thus requires a body, angels cannot be temperate; but it is better to be an angel than a human being). Anselm concludes that only those terms of the first type, those by which it is always better to be named than not, are of the sort that can be purified sufficiently to name God’s essence. These are the pure perfections. Anselm gives several lists of these, the longest of which is at the end of Monologion 16:31


---

31 (H&R): Illa igitur est summa essentia, summa vita, summa ratio, summa salus, summa iustitia, summa sapientia, summa veritas, summa bonitas, summa magnitudo, summa pulchritudo, summa immortalitas, summa incorruptibilitas, summa immutabilitas, summa beatitudo, summa æternitas, summa potestas, summa unitas, quod non est aliud quam summe ens, summe vivens, et alia similiter.
There remain a few difficulties with predicating these perfections of God. If God is to have such qualities as wisdom or beauty, from where does He get them? He could not have them from outside Himself, since that would imply that these qualities existed independently of God and that God had these qualities “by participation” (*participazione*) in them. But this is not true; these qualities are not like independently existing Platonic Forms, in which God must participate. Therefore, all things that God is He is through Himself (*per se*), and all things that are exist through God. If God is just or wise, He is just or wise *per se*. Since He is just through Himself, He is the cause of His own justice; since justice is defined as that by which just things are just, it follows that God Himself is nothing other than justice. A similar path of reasoning can lead to similar conclusions about the other perfect qualities, so that God is nothing other than wisdom, beauty, and so on; He is each of these things through Himself, He is each of these things supremely, and He is each of these things essentially.

---

32 M, ch. 16.

33 Cf. M, ch. 3.

34 M, ch. 16. The preceding paragraph has been a quick summary of the argument of ch. 16. Anselm’s unusual use of ‘*participazione*’ seems to follow Augustine’s use in *De Trinitate* 5.10.11, in a similar context. It might also suggest that here he is taking up the central idea of Boethius’s *De Hebdomadibus*, in which Boethius introduces Aristotelian metaphysical and logical distinctions to trim away the excessive ontological commitments of the Neoplatonic idea of participation.

Surprisingly, this connection with the *De Hebdomadibus* does not seem to have been even considered before: Desmond Paul Henry does mention this passage of M 16 when he argues against Ockham’s appropriation of this text in *Summa Logicae I.7*, but he seems to think that Anselm is being original (as does Ockham) (*The Logic of St. Anselm* (Oxford University Press, 1967), §2.22). He also suggests that Anselm derived key premises for his *Proslogion* argument from the *De Hebdomadibus*, but he seems to leave open the possibility that Anselm derived it from a commentary on Aristotle by Pope Sylvester II (938-1003). Gillian Evans denies that there is any evidence that Anselm even
Furthermore, the non-relational names that can be predicated of God—e.g. ‘wise’, ‘beautiful, ‘just’, and so on—are, when said of creatures, predicated accidentally. They are said to be in or of the individual creature, so that, unlike with relations, the creature really changes if the quality changes. If an object were white and were to become blue, there would be a change in the object. If qualities were to be predicated of God, then, since qualities are accidents, it would be conceivable that God could remain what He is without being wise, just, or beautiful. The purification procedure must then be applied not only to what these terms signify, but also to how they signify, so that they do not introduce into the concept any suggestion that it indicates an accident.

read Boethius’s theological tractates (Anselm and Talking about God (Oxford University Press, 1978), p. 45). This seems wrong, not only because of M 16, but also because M 79 uses Boethius’ definition of ‘persona’ from Contra Eutychen 3.4-5.

35 This does is not true of relational accidents, though. If an object were greater in size than the object next to it, and the second object were altered so that it became larger than the first, there would not be a change in the first object, even though it could no longer be considered greater than its neighbor. But only accidents of relation can be handled this way.

36 M, ch. 25.

37 Aquinas makes a distinction between what a name signifies and its mode of signification (STh Ia.13.3), arguing that divine names take what they signify from God and their mode of signification from creatures, and are therefore analogical names. Anselm is wrestling with the same problem, but I think he would find this distinction too slippery—he thinks the way ‘wisdom’ signifies also affects what it signifies, so that to understand God properly, one must purify one’s verbum of all limitations that derive from human ways of knowing. In the case of wisdom, one must remove not only the idea that wisdom is the ability to teach hard-won lessons about life (because that is not what God’s wisdom is), but also the idea that wisdom is necessarily an accident (since in God it is not). Aquinas’s distinction enables one to discuss God’s attributes scientifically, because it demarcates the limits of human comprehension in such a way that it enables rational analysis. Anselm’s goal is an ascetical and mystical one—he wants to understand God as much as he can in this life, and so he would be reluctant to accept Aquinas’s distinction if it led people to be satisfied with less-than-pure modes of signification.
It might seem that if each of the perfections were predicated of God’s substance, then that substance would be composed of several perfections. But whatever is composed of parts requires its parts in order for it to exist; for instance, a car could not exist if its tires, body, and engine did not also exist. If God were composed of parts, His existence would be dependent on these parts, and so He would not exist per se. By the same argument used to show that the perfections did not exist independently of God, it would appear that God could not be composed of constituents, even such perfect constituents as justice; therefore God is wholly simple. Anselm concludes:

Therefore, since this Nature is in no way composite and yet is in every respect all those good things, it follows that all those good things are not many, but one. Hence, any one of them is the same as all the others—whether they be considered together or separately. For example, when [this Nature] is called justice or being, these predicates signify the same thing as the other predicates (whether taken together or separately). Thus, just as whatever is predicated essentially of the Supreme Substance is one, so whatever the Supreme Substance is essentially, it is in one and the same way and respect.

All the perfect attributes, all those things whose names can be predicated of God’s substance because they are better to be than not to be, which are attributed to God essentially and not accidentally, are one. This is a remarkable conclusion, and yet it makes a certain sense. God cannot be other than His perfections, since that would violate His simplicity and thus His independence; yet He cannot be missing a single perfection

---

38M, ch. 17; H&R 1:27: Cum igitur illa natura nullo modo composita sit, et tamen omnimodo tot illa bona sit, necesse est ut illa omnia non plura sed unum sint. Idem igitur est quodlibet unum eorum quod omnia, sive simul, sive singula. Ut cum dicitur iustitia vel essentia, idem significat quod alia, vel omnia simul vel singula. Quemadmodum, itaque unum est quidquid essentialiter de summa substantia dicitur, ita ipsa una modo, una consideratione est quidquid est essentialiter.
and be perfect Himself. Since these attributes do not imply any limitation of perfection, they can all be attributed to God without implying any limitation or composition.\(^{39}\)

At this point, however, it seems that Anselm has gone too far, for at least two reasons. First, it is difficult to swallow that beauty, substance, justice, and wisdom are identical. Second, if God’s attributes are nothing other than His nature, it seems that the attempt to understand God’s attributes, which was supposed to be more manageable than trying to comprehend His nature, has collapsed back into the problem of understanding God’s nature. Let’s look at these in order.

Anselm is committed to saying that the pure perfections have to be capable of being identical with the divine nature. This is counterintuitive, at least. But it seems wrong on its face if we consider Anselm’s other claim that the pure perfections are also to be applicable to creatures. Even though it is better to be substantial, just, and wise, than not to be these things, and thus it can be shown that these things are capable of being identical in God, it is not exactly clear that all good desserts are just, or that all beautiful people are wise. ‘Beautiful’, ‘just’, and ‘wise’ all have different definitions; they are each understood by a different *verbum*; and they certainly don’t seem to be the identical in creatures and thus in themselves.

We can solve part of this problem rather quickly. As Anselm argues in M 17, our concept of each perfection will be purified even further by the constraints of the unity of all perfections. Presumably, Anselm anticipates us holding all our concepts of the divine

---

\(^{39}\) For example, if the Supreme Nature were described as “flesh and bone”, it would seem to imply that it was composed of those two material substances and that it was limited in space and time. Describing the Supreme Nature as “wise and just” lacks these connotations.
attributes in something like what Rawls calls ‘reflective equilibrium’, where the requirement that all divine attributes are the same leads us to go back and forth among them, removing more and more of the vestiges of their creaturely origins, where those are inconsistent with the other attributes and with what we can understand of the divine nature. This just adds a further stage to the purification procedure, and thus does not create any additional theoretical obstacles to our account of how it is possible to understand the divine nature.

It does suggest, though, that Anselm ought to be committed to a very strong view of the unity of the pure perfections in creatures, so that creaturely wisdom would have to indicate a sort of beauty and creaturely justice a sort of goodness. The attributes in creatures would not be pure as such, since ‘creaturely attribute’ cannot describe a perfection whose concept is purified of all references to creatures. The purification procedure cannot remove from our concepts of the perfections of creatures that they are accidents, for example. But the concepts when applied to creatures will be purified of all other features—references to the body, e.g.—that are specific to creatures. We will see this in chapter eight when we try to apply Anselm’s account of divine justice to the conditions of human justice, and have to take into account that where God is omniscient and omnipotent, human beings are ignorant and relatively impotent.

The second problem is trickier. It seems that all Anselm’s work trying to develop definitions of the attributes that can apply equally to God and man seems to have been futile, since God’s simplicity has made His attributes unable to be understood. In the next section, we will suggest how Anselm responds to this problem about understanding the attributes.
5.7 Are Pure Perfections Really Any Help?

So far, we have described three possible categories of human concepts: thoughts of creatures only, thoughts that fully comprehend God’s nature, and an incomplete yet not false grasp of the attributes of God through purified concepts derived from appropriate categories of creaturely attributes. The current worry before us is that the third position is just as impossible for humans as the second, since our creaturely nature limits what we are able to conceive.

Let us divide these creaturely limits into two groups. The first group we have already discussed: the epistemological limits that arise from trying to describe an infinite God with concepts derived from finite creatures. But even were we to derive our concepts directly from God, there would be still be limits on our comprehension of Him. When the saints in heaven look on God face-to-face, their concepts of God are derived from God himself rather from the faint images found in creatures, yet they still fail to understand God completely because they remain lesser beings with minds inadequate to the task.

Human beings can fail to comprehend God for what we might call ontological as well as epistemological reasons. The ontological obstacles are obviously permanent, but Anselm hopes the epistemological ones might not be.

But if we are to overcome the epistemological limits, it seems that we must answer the question of how it is humanly possible to develop a verbum that signifies God. Anselm discusses a similar problem in De Casu Diaboli 11 when he analyzes the signification of negations and privations, e.g. ‘nothing’, ‘not-man’, ‘evil’, and ‘blindness’. I say the problem is similar because it is quite hard to imagine how one can have a verbum of nothing, especially when we remember that Anselm’s idea of the
verbūm was explained by appealing to the visualization of a building that an architect has in his mind before he builds it. It is not clear that visualizing nothing is much easier than visualizing divine simplicity; yet Anselm thinks it possible to form some sort of verbūm of both nothingness and God.

D. P. Henry points out \(^4\) that the modern way of handling negative terms such as ‘not-man’ is with quantifiers, as in “for any \(x\), \(x\) is a not-man if there is a \(y\) such that \(y=x\) and \(y\) is not a man.” As Henry also points out,\(^4\) Anselm was aware of this sort of solution and makes use of it in M 19. But he uses an analysis of an entirely different character when he talks about ‘evil’ and ‘nothing’ in DCD 11. Since he knows the modern-style solution, he must reject its use here for a reason.\(^4\) Here’s Anselm’s argument:\(^4\)

\(^{40}\) Henry, The Logic of St. Anselm, §6.64.

\(^{41}\) Ibid., §6.66

\(^{42}\) I think the part of the difference between Anselm’s solution and the modern one goes back to his idea of the verbūm. The tendency of modern logicians is to design their formal languages the way programmers try to design platform-independent software. That is, they try to develop languages that are compatible with just about every ontology or theory of the mind, just as programmers develop programs that can run on both Windows and Macintosh operating systems. This is an admirable goal, and ideally it should allow philosophers to compare their theories in a language neutral between them. Yet Anselm did not have this goal, and often his claims about words and concepts are heavily dependent on his theories about the intellect. Like platform-dependent software, this has benefits and costs: there is greater integration between the various theories, but one pretty much has to buy the whole package or else things don’t work as well. Richard Rorty’s pragmatism is a contemporary example of platform-dependent philosophizing: one has to accept the whole pragmatist project or else one is guilty of holding on to useless vocabularies and outdated distinctions.

\(^{43}\) (With slight changes from H&R): M. Constat quoniam haec vox, scilicet “nihil,” quantum ad significationem nullatenus differret ab eo quod dico “non-aliquid”. Nihil quoque hoc apertius, quam quod haec vox, scilicet “non-aliquid” omnem rem penitus et omne quod est aliquid intellectu removendum, nec omnino ullam rem aut penitus quod aliquid sit in intellectu retinendum sua significatione constituit. Sed quoniam remotio
T. It is evident that this word, viz., “nothing,” does not at all differ, with respect to its signification, from what I term “not-something.” Also, nothing is clearer than that the word “not-something” indicates by its signification that absolutely every thing and all that is something should be removed from the understanding, and that no thing whatsoever nor what is at all something should be retained in the understanding. But the removal of a thing cannot at all be signified except together with the signification of that very thing whose removal is signified. (For example, no one understands what “not-man” signifies except by understanding what a man is.) Therefore, it is necessary that the word “not-something” signify something by destroying, so to speak, that which is something. But since by removing everything that is something, the word “not-something” signifies no being which it indicates is to be retained in the understanding of the hearer, it signifies no thing nor what is something.

Therefore, by means of these different considerations, the word “not-something” does in some respect signify a thing and something, and does not in any respect signify a thing or something. For it signifies by removing (removendo) and does not signify by establishing (constituendo). In this manner, the name “nothing,” which destroys everything that is something, signifies something rather than nothing by destroying and signifies nothing rather than something by establishing [i.e., by positing]. Therefore, it is not necessary that nothing be something simply because its name somehow or other signifies something. Rather, it is necessary that nothing be nothing, because its name signifies something in the aforementioned way. And so, in this aforementioned way the

alicuius rei significari nullatenus potest nisi cum significatione eius ipsius cuius significatur remotio -- nullus enim intelligit quid significet “non-homo” nisi intelligendo quid sit homo --: necesse est ut haec vox quae est “non-aliuid,” destruendo id quod est aliquid significet aliquid. Quoniam vero auferendo omne quod est aliquid, nullam significat essentiam quam in audientis intellectu retinendam constituet: idcirco “non-aliuid” vox nullam rem aut quod sit aliquid significat.

Igitur haec vox “non-aliuid” his diversis rationibus aliquatenuis significat rem et aliquid, et nullatenus significat rem aut aliquid. Significat enim removendo, et non significat constitutendo. Hac ratione “nihil” nomen, quod perimit omne quod est aliquid: et destruendo non significat nihil sed aliquid, et constitutendo non significat aliquid sed nihil. Quapropter non est necesse nihil esse aliquid, ideo quia nomen eius significat aliquid quolibet modo; sed potius necesse est nihil esse nihil, quia nomen eius significat aliquid hoc modo. Hoc itaque modo non repugnat malum nihil esse et “mali” nomen esse significativum, si sic aliquid perimendo significat ut nullius rei sit constitutivum.
fact that evil is nothing is not opposed to the fact that the name “evil” is significative—provided that “evil” signifies something by destroying [i.e., by negating] it and, thus, is constitutive of nothing.

We think of blindness by thinking of something which is able to see, and then thinking of the thing without its power to see (a blind thing), and then removing the explicit reference to the subject (blindness). So blindness signifies the absence of sight in any thing that ought to have sight. We could carry this via removendo further, by removing from our concept of blindness the suggestion that the thing ought to have the power of sight; we would then have the negative concept of non-sightedness, as in “a stone is non-sighted.” The negative concept nothing can be formed in a similar way: when you tell me that there is nothing in the cupboard, I don’t just think of the cupboard itself, as I might if you had just brought it home from the store and were pointing out its various features. I think of the cupboard as empty, as missing something that it might or should have. If we take the idea of empty cupboard and remove from it the idea of cupboard, we have the privative idea of empty space, some bounded area that is missing something. If we subtract the idea of lacking something, we have the idea of space. If we subtract from space the idea of boundary, we have the negative concept nothing.44

Anselm appears to give a uniform treatment of negative names such as ‘nothing’ and privative names such as ‘blindness’. Henry (Logic §6.652) thinks this is a confusion, and suggests that Anselm overlooked Aquinas’s point (STh Ia 48.3 ad 2) that a negation does not need to be in a subject while a privation does. I think that Henry misses Anselm’s purpose in the passage of DCD under discussion, which is to explain how we can form a mental word of a not-something. One can make Aquinas’s distinction between privations and negations and still acknowledge that the two sorts of concepts are formed through the same via removendo. I think Anselm has the resources to make Aquinas’s distinction; cf. DCD 12, where he introduces the idea of unactualized potency, which is
The purification procedure is different from a privation or a negation in that it removes from the concept an imperfection or limit rather than a perfection. Indeed, it is this difference that makes the purified concept more difficult to grasp than most negations or privations. In the case of an ordinary privation such as blindness, we are familiar with the perfection that is being removed (sight), and we are familiar with the being when it contains the perfections it should have (a person with normal vision). In the case of the purification procedure, however, we are familiar only with the limit or imperfection that is being removed, and we don’t usually have any experience of something not limited in that way. (One might argue that the concept ‘nothing’ is hard to conceive for a similar reason, since it also signifies a lack of limits.)

Anselm’s distinction between what is signified *removendo* and what is signified *constituendo* is helpful for our purposes. The word ‘nothing’ signifies ‘something’ as the original concept from which it was derived by the *via removendo*; it does not signify ‘thing’ in any positive sense. The same goes for ‘evil’, which remotively signifies ‘good’ but constitutively signifies nothing. I think that Anselm’s purification procedure for divine attributes is a specialized application of the *via removendo*, and that the distinction between remotive and constitutive signification allows us to solve the problem posed at the end of the last section. In negations and privations, the word signifies only remotively and not constitutively. In divine names, both sorts of signification are present. The original creaturely concept is signified remotively (i.e., it leaves its trace on our thoughts), while the pure perfection is conceived constitutively (i.e., it provides the itself a privation. Nonetheless, the act of concept formation is similar for privations and negations, as has just been shown.
positive content of the concept). Thus, after purification, ‘wisdom’ signifies the wisdom of Socrates, Plato, Aristotle remotely, and the divine wisdom constitutively.\(^{45}\)

So here’s Anselm’s theory. We begin with a good quality of creatures P. We ask whether it is always better to be P than not to be P. If not, then we remove one imperfection from the *verbum*. If the removal of that imperfection would introduce a contradiction into the definition of P, we have to pick another quality. If not, then we ask whether it is always better to be P\(^*\) than not (where P\(^*\)=P – the imperfection). If not, we remove another imperfection and repeat. If so, then we have a pure perfection, one that can apply equally well to God and creatures. When we have several pure perfections, we ask whether these are consistent with each other. If not, then we use the purification procedure in a process of reflective equilibrium to remove the inconsistencies.

We are now able to consider how Anselm applied these conceptual tools to divine justice and mercy. We will begin by looking at an early text, his prayer to St. Paul, where Anselm first questions the juridical framework of the atonement. From there we will consider what he has to say about justice and mercy in his non-devotional works, starting with the *Monologion* and working in chronological order through the relevant parts of the entire corpus. In doing so we will see that justice is a central concern of Anselm’s

\(^{45}\) Each complex concept contains its history, so to speak; the act of conceptual subtraction is not arithmetical, as if “sight – sight = 0.” Instead it is like erasure, where the vestiges of what was erased from the more basic concept affect the signification of the more complex concept derived from it. I think that this contrast between the ‘arithmetical’ and ‘historical’ modes of concept formation suggests why most modern formalizations of ‘nothing’ do not treat it as a name but instead as a shorthand for “some arrangement of quantifiers,” in Henry’s phrase (*Logic* §6.641). I am unaware of any formal language that allows the history of a concept to affect its signification, although as we have seen one can develop a psychologically plausible account of concept formation where this happens.
thought, and his intellectual efforts after the publication of the *Proslogion* were driven by his desire to develop a pure concept of justice that could allow him to have some contact with the nature of God.
CHAPTER SIX

ST. ANSELM AS A THEORIST OF JUSTICE

6.1 The Development of Justice and Mercy in Anselm’s Speculative Works

Few people seem to have noticed or made much of the fact that justice is a persistent theme in nearly all of Anselm’s speculative writings.\(^1\) Anselm is famous for his theory of the atonement, for his method of philosophical theology, and for the argument for the existence of God in the *Proslogion*, but he is not often thought to have been a political thinker.\(^2\) As we will show in the next chapter, recent historical scholarship is starting to show that Anselm was politically successful, and so it is no longer surprising to think that he might have turned to political matters in his speculative thought. Philosophers and theologians have not really examined this historical scholarship, though. As a result, it is still commonly argued that Anselm borrowed his ideas of justice

\(^1\) For example, Richard Southern (*Saint Anselm: A Portrait in a Landscape* (Cambridge University Press, 1990)) mentions justice only in the context of DV 12 (p. 172) and in the CDH (213-6), while Jasper Hopkins (*A Companion to the Study of St. Anselm* (University of Minnesota Press, 1972) has no significant treatment of justice in the entire book.

\(^2\) Southern writes, “law and politics had no place in his vocabulary…much in his later life as archbishop becomes clearer if we remember that every practical question had for him a monastic orientation” (*Saint Anselm*, p. 172).
from the culture of his day. As we will show in this chapter, Anselm wrote about justice and themes relating to justice in almost all his major works, and his views about justice, rather than being flatly conventional, are quite original, for his day and ours.

In this chapter, we’ll take up his works in chronological order, starting with his prayer to St. Paul, and then looking at the *Monologion* and *Proslogion* (written while he was prior of the abbey at Bec), the three dialogues *On Truth*, *On Free Choice of the Will*, and *On the Fall of the Devil* (written while abbot at Bec), and the *Cur Deus Homo* (written while archbishop of Canterbury). As we go along, we’ll see how each work contributes some small piece to the complex view that Anselm deploys to such effect in *Cur Deus Homo*.

### 6.2 Anselm on Justice and Mercy in the *Oratio ad Sanctum Paulem*

Anselm begins *Oratio* 10 with a juridical image—because of his sins, he is accused by everyone, defended by nobody, and is most terrified of the strict judge God. According to justice (*secundum justitiam*), he deserves to be damned. He explores why he is so abject—his sins (*crimines*), his inability to hope, his likelihood of damnation: “Since you voluntarily made yourself wretched, it is necessary by justice that you be

---

3 Even as astute an Anselm scholar as Marilyn McCord Adams assumes that the concept of justice for Anselm is tied up with an honor and shame culture. See *Horrendous Evils and the Goodness of God* (Cornell University Press, 1999) ch. 6.

4 Anselm probably included this among several prayers sent to Adelaide, a daughter of William the Conqueror, in 1072. See Southern, *Saint Anselm*, pp. 91-93.
wretched always.”5 Then he appeals to the goodness (bonitas) and mercy (misericordia) of the kind God (benigne domine) to give him legal counsel how to get out of this situation. This is a strange moment for the legal metaphor—the criminal appeals to the judge for legal advice for how to get off the hook! Then he appeals to God by name, pointing out that Jesus’ incarnation has no purpose except to save sinners such as Anselm. Clothing himself in faith, he asks God to hide him from the judgment of God.

He shifts away from legal metaphors in the next section of the prayer. He is unable to hide behind the cloak of faith, because he realizes that his sinful actions betray his lack of faith (oddly to our post-Reformation ears, in a prayer to St. Paul, the source for Luther’s doctrine of sola fide, Anselm quotes St. James’ admonition that “faith without works is dead”). “Those who are sterile in good are dead,” he realizes; not completely dead, but certainly heading toward death and certain damnation, because they are unable to raise their dead spirits to offer anything to God. So Anselm realizes that he is worse off than he thought he was before: he is not just guilty, he is no longer spiritually alive. So he appeals to St. Paul, who said that the grace of God was sufficient for him to do all things (2 Cor 12:9). These words give the dead man hope, for surely if he can do all things he can intercede on behalf of the penitent sinner, and can even raise the dead. Anselm treats the Pauline verse as a promise which must be kept, echoing his point earlier in the prayer that the incarnation of Christ requires the salvation of sinners—the power and mercy of the saint and the savior creating an obligation to aid the soul in need.

5 Oratio 10 (translation from The Prayers and Meditations of Saint Anselm, trans. Benedicta Ward; henceforth, BW): Utique qui te sponte fecisti miserum, iuste necesse est semper te esse miserum.
Indeed, he claims of Jesus and Paul that “they seek to be importuned by the suffering, they wish the wretched to be persevering in prayer, they love those who lament unceasingly,” as if they needed to show mercy to be fulfilled.

The final section is even more surprising. Anselm declares that because of their tender mercy, Paul and Jesus are mothers and nurses. Since they are also just, they are fathers as well: 6

You are fathers by your effect, and mothers by your affection; fathers by your authority, mothers by your kindness; fathers by your teaching, mothers by your mercy.

We saw earlier how at the beginning of the oratio Anselm deliberately uses juridical language (“accuser”, “crimes”, “judge”, “counsel”, “strictness”) in order to undermine the legal understanding of sin. There we saw him break down the logic of the juridical metaphor by appealing to God’s mercy and kindness to preserve him from the judgment of God’s justice and strictness. Here we see the same juxtaposition of seeming contraries in a more daring way, by attributing femininity to the masculine Paul and Jesus. We can see in this a foreshadowing of the later theme that justice and mercy are identified in God, a poetical prelude to the reflective equilibrium stage of the purification procedure.

In the very next passage Anselm introduces another theme that he develops in his later works. While Jesus and Paul are both mothers, the one is God and the other is not: 7

6 Oratio 10 (BW): Patres igitur estis per effectum, matres per affectum. Patres per auctoritatem, matres per benignitatem. Patres per tuitionem, matres per miserationem.

7 Ibid. Etsi quantitate affectus impares, in qualitate tamen non dissimiles. Quamvis magnitudine benignitatis non coaequantes, voluntate tamen concordantes. Licet plenitudine miserationis non convenientes, intentione tamen non disadvenciontes.
If in quantity of affection you are not peers, yet in quality you are not unalike; though in the greatness of your kindness you are not equals, there is concord between your wills; although in the fullness of mercy you do not agree, yet in intention you are not disagreeing.

The emphasis here is on the similarity between God and man across the transcendental divide. As we have already seen, Anselm refines his theory of divine names from this poetical beginning; nonetheless, it is significant that even in this early text Anselm is holding that the mercy of God and of the saints is in the most significant respects the same.

Anselm begins with a juridical metaphor, which he subverts in a way that seems surprisingly postmodern. He ends with the very different image of a mother, whose fundamental attitude is affection and caring, who cannot but respond to the cries of her child, who embraces and heals and comforts the sinner’s wounded soul. Anselm is often accused of having too juridical a picture of God, but we will see that the themes of this oratio are closer to his real views. Even though God is demanding, and must be so from His just nature, at the end he should be seen as affectionate and merciful.

6.3 Justice in the *Monologion*\(^8\)

Although few people seem to have noticed this, justice is an important topic in the *Monologion*.\(^9\) To take a superficial example, justice is one of Anselm’s favorite examples

\[^8\] Southern dates the *Monologion* to around 1075. Cf., *Saint Anselm*, p. xxvii.

\[^9\] E.g., Marilyn Adams, in one of the few articles covering major themes in the *Monologion* since the critical edition of Anselm’s works was published (“Romancing the Good: God and the Self according to St. Anselm of Canterbury” in *The Augustinian Tradition*, ed. Gareth B. Matthews (University of California Press, 1999), discusses
of an attribute of God—God’s Justice and Goodness appear throughout the text. For most of the *Monologion*, Anselm doesn’t treat justice differently than the other divine attributes, but he frequently uses it as an illustration of some more general theoretical point. So in chapter 1, Anselm argues that all just things, to whatever extent they are just, are so through justice, and that justice is the same in each case, if not in degree at least in kind. He uses this point to argue that all good things are good through one goodness, which, he claims, must be existent goodness itself. In chapter 15, he claims that it is always better to be just than unjust, so that justice is a candidate for being a pure perfection and a divine attribute.

More importantly, the last fifteen chapters of the *Monologion* are taken up with an argument from God’s justice to the immortality of the human soul and the necessity of faith. Anselm’s mature reflections on justice and mercy are found in the *Cur Deus Homo*, completed nearly two decades later. But these neglected final chapters of the *Monologion* allow us to plot a trajectory of his thoughts on justice. Let’s take a look at the *Monologion*’s three substantive claims about Divine Justice:

1. God’s Justice *necessitates* (a) human immortality, (b) the eternal punishment of the wicked, and (c) the eternal beatitude of the just (M 71-2);
2. Since these facts about the life to come are accessible to human reason, it is unreasonable and unnatural for a rational being to lack faith, hope, and love of God (M 75-6); and
3. Since God is Justice itself, the rational creature who loves God should desire to bring about a just order wherever he can. “Indeed, this faith, … is not at all idle—provided the opportunity to use [it] arises. Rather, [this faith] exercises itself in a great number of works…. [These claims] can be proved by the solitary fact that what loves Supreme Justice can neither

Anselm’s account in the *Monologion* of how the soul can reach God without mentioning justice.
This third point is quite important for Anselm’s political theory—the love of Supreme Justice requires that one love lesser forms of justice as well, and it requires that one desire to effect justice wherever possible. Anselm concludes from this (very Platonic) argument that faith without works is dead.\footnote{10}

It might be helpful to look at these claims in more depth. In chapter 64, Anselm takes a step back from his discussion of the divine nature to reflect on faith and reason: What we can say about the divine nature? What we can think about the divine nature? How interesting is it that we cannot think everything that we can in confidence say? For the rest of the treatise, he focuses on the relation of human intellect and will to the divine essence, and he wants to argue, without appealing to Scripture or other authorities, that if one loves God and does good works for His sake, one will live in eternal beatitude.

The argument for the immortality of the soul, which properly begins in M 66, is interesting for what it shows us about his views of justice at this stage. Since it is impossible to comprehend God directly, Anselm reasons, one can only know Him through that which is most similar to Him, the rational mind (\textit{mens rationalis}). Since the mind is able to love, understand, and be conscious of itself, its operation is similar to the

\footnote{10} (M 78; H&R translation). \textit{Etenim nullatenus fidem illam quam competens comitatur dilectio, si se opportunitas conferat operandi, otiosam esse sed magna se quadam operum exercere frequentia, quod sine dilectione facere non posset, vel hoc solo probari potest, quia quod summam iustitiam diligite, nihil iustum contemnere, nihil valet iniustum admittere.}

\footnote{11} Anselm also emphasizes this text from St. James in his \textit{Oratio to St. Paul}, which Southern dates as before 1070.
activity Anselm has earlier attributed to the three persons of the Trinity. Also, since the activity of the mind involves the mind to some degree becoming like the thing known, the rational mind will be maximally similar to God when striving to love, understand, and be conscious of God.

Since God has created man with this ability to become a better image of the Supreme Essence, Anselm concludes that man owes it to God to strive to actualize this ability. This follows for three reasons: First, every creature owes God everything that it is, and so it ought to give everything back to God, with interest. Second, even placing that duty aside, one ought to want to do the greatest thing within one’s power, which in this case is to become like God. Finally, even placing aside that natural desire, the activity of rationality involves the making of distinctions between what is just and unjust, between what is the greater and lesser good, and ordering one’s likes and dislikes accordingly; thus it is essential to the nature of rational beings that one love the Supreme Good above all other goods and to love other things only on account of that which is Goodness itself. And love follows upon awareness and understanding of the loved. So the only way a rational being can fulfill its nature is to do everything it can to understand, love, and be aware of God; indeed, it must recognize that such is the purpose of its creation.

But God, having created rational beings with this end, has committed Himself in justice to sustain in perpetuity those rational beings that do strive to love Him. For it

\[\text{12 M 29-63.}\]

\[\text{13 As we saw in M 10, the\ verbum\ is some sort of similitude of the thing known.}\]

\[\text{14 M 68.}\]
would be cruel to create rational beings so that they love God but can only actualize that love for a limited time, after which they either turn against God or are involuntarily forced to stop loving Him. But it is absurd to think that God would have created a rational being that would understand the greatest good and contemn it, or, Anselm claims, to violently force a creature to stop loving Him whom it was created to love. God therefore created the soul to love Him without ceasing; it follows that it must live without ceasing, at least so long as it loves Him. What’s more, Supreme Justice would not allow such love to go unrequited, so it must give the rational being what it desires—itself. So it follows from the justice of God, that “every soul which strives, as it ought, to love and desire supreme blessedness will at some point partake in its delights.” Anselm in subsequent chapters argues that divine justice also necessitates that the souls which voluntarily reject God suffer eternal wretchedness, and that the souls of infants, which

15 M 69.

16 M 70: Nihil ergo verius, quam quod omnis anima rationalis, si quemadmodum debet studeat amando desiderare summam beatitudinem, aliquando illam ad fruendum percipiat.

17 M 71. I don’t contest Anselm’s conclusion here, but the argument he gives is particularly bad, and is inconsistent with principles Anselm holds elsewhere in his writings. Anselm argues that the souls which reject God should be punished either with eternal wretchedness or with annihilation; but annihilation would return the souls to the same state as the ‘innocent’ souls which are not yet created, and so would not really be a punishment for they would be as if they were not yet born. Since that would be unjust, the evil souls must be immortal and punished forever. But for Anselm, famously, it is always better to be than not to be, which means that it is better to be and be wretched than not to be. I think Anselm has the better argument in CDH II.16, where he appeals to God’s mercy as the grounds for the immortality of even the wretched—that He could have annihilated them, but instead allows them to remain in the existence of their own choosing. But mercy does not play a significant role in the Monologion.
neither accept nor reject God, also should be immortal (because they share human nature which is immortal).\textsuperscript{18}

Anselm’s last statements on justice in the \textit{Monologion} explain why justice demands of rational beings that they perform good works as well as that they love and believe in God. The argument is quite simple: to love Supreme Justice is to reject nothing just and to accept no injustice, since all just things are just by resemblance to the Supreme Justice, and all things similar to the Supreme Justice are deserving of love because of that similarity. So the soul in love with God will, given the ability and opportunity, overflow with works of justice and works opposing injustice.\textsuperscript{19}

These last chapters of the \textit{Monologion} are important for our understanding of Anselm’s reflections on justice. Anselm thinks his philosophical work earlier in the treatise allows him to make substantive claims about divine justice: that it necessitates human immortality, the eternal punishment of the wicked, the eternal beatitude of the just; since these facts about the life to come are accessible to human reason, it is unreasonable and unnatural for a rational being to lack faith, hope, and love of God; and since God’s intention for each rational creature is that it love God, it follows that the rational creature should have an equal desire to bring about a just order wherever he can. He thinks his arguments allow him to probe more deeply into the logic of Christian holiness. Whether or not this is the case, the attempt forces him to say more about divine justice than that it is merely another ineffable attribute of God. Whatever we might think

\textsuperscript{18} M 72

\textsuperscript{19} M 78
about his actual arguments here—and some of them are quite tendentious—we can mark this stage of the treatise as the beginning of Anselm’s deep and lifelong attempt to fathom the actual content of divine justice by comparing it to human justice. In the Proslogion, to which we turn next, he covers some of the same territory, but seems almost to stumble into his first discussion of God’s mercy and its relationship to divine justice.

6.4 Goodness, Mercy, and Justice: Navigating the Proslogion

In the entire Monologion, there is not one mention of the mercy of God—the word misericordia doesn’t even appear in the text. The account of God’s justice at the end of the Monologion is straightforward: the good are rewarded with eternal bliss, and the bad with eternal punishment. One can imagine Anselm’s students getting a little worried when they read this, because according to the doctrine of original sin as well as their own experience, all people are sinners. If Anselm’s arguments were the last word, then all men would be damned. In response to this problem, in the very first chapter of the Proslogion Anselm begs God for mercy, and in chapters 8-11 he for the first time examines the mercy of God philosophically.

In Proslogion 6, Anselm produces a surprising list of attributes that he claims are pure perfections—“sensation, omnipotence, mercy, and impassibility”20—and in the subsequent chapters he sets out to prove how they are so and are one in God. This list is surprising because it includes two relatively unproblematic divine attributes, omnipotence and impassibility, amidst two attributes, sensation and mercy, which seem quite

20 P 6: Verum cum melius sit esse sensibilem, omnipotentem, misericordem, impassibilem quam non esse.
problematic. Before we look at the Anselm’s solution to these problems, we might wonder about his motivation in including them at all.

Two textual clues from the Prosligion, where Anselm uses miserare and sensibilis in contexts that we might call more rhetorical than philosophical, suggest an answer. In P 17, Anselm wants to say that God is beautiful, harmonious, fragrant, sweet, and soft, because such a God would be most pleasing to the five senses, and so he argues that God has in his ineffable way what he has given to his creatures in a sensible way. That God is ineffably sensible (sensibilis) does not exactly explain why Anselm thinks God has sensations, but it suggests an explanation—if God is in some way sensible, and he is to know himself completely, then he would have to be in some way sensitive. Anselm does conclude that sense perception is a sort of knowledge, and since God is all-knowing, he is supremely sensitive, so this conjecture has some plausibility. In any case, it does seem that his poetic imagination is driving his philosophical conclusions.

We can conjecture a little more confidently about the inclusion of mercy among God’s attributes. At the end of his treatment of mercy in chapter 8, Anselm argues, “if you are not merciful, whence comes so much consolation for the sorrowful?” This poignant worry harkens the reader back to the first occurrence of miserare in the

\[\text{\underline{21}}\] Clearly Anselm is less stringent here than in M 15—if God can be sweet in some ineffable way, why can’t he be gold in some ineffable way as well? Presumably, the difference in what counts as a candidate for divine attribution is due to the different rhetorical context of the Prosligion.
Proslogion’s first chapter, entitled “A Rousing of the Mind to the Contemplation of God”: 22

O Lord, how long? How long, O Lord, will You forget us? How long will You turn away Your face from us? When will You look upon us and hear us? When will You enlighten our eyes and show us Your face? When will You restore Yourself to us? Look upon us, O Lord; hear us, enlighten us, reveal Yourself unto us. Restore unto us Yourself—without whom we fare so badly—so that we may fare well. Have mercy upon the efforts and attempts which we, who can do nothing without You, direct toward You. [As] You summon us, [so] aid us, I beseech [You], O Lord, that I may not despair with sighing but may revive in hoping. I beseech You, O Lord: my heart is made bitter by its own desolation; sweeten it by Your consolation. I beseech You, O Lord, that having begun in hunger to seek You, I may not finish without partaking of You. I set out famished; let me not return still unfed. I came as one who is poor to one who is rich, as one who is unhappy to one who is merciful; let me not return empty and spurned.

Such gloomy language is a stark contrast to the optimistic last chapters of the Monologion, where Anselm proved that all rational beings are by nature destined to spend eternity with God so long as they act rationally. In M 70, Anselm concludes that divine justice requires that in the life to come all successfully rational beings will see God face to face; now, Anselm is upset that God has turned his face away. There he was the rationalist philosopher, here he is the desperate lover. At the end of P 1, he recalls his

22 P 1 (H&R) Et o »tu, domine, usquequo«? »Usquequo, domine, oblivisceris« nos, »usquequo avertis faciem tuam« a nobis? Quando respicies et exaudies nos? Quando illuminabis oculos nostros, et ostendes nobis »faciem tuam«? Quando restitues te nobis? Respice, domine, exaudi, illumina nos, ostende nobis teipsum. Restitue te nobis, ut bene sit nobis, sine quo tam male est nobis. Miserare labores et conatus nostros ad te, qui nihil valemus sine te. Invitas nos, »adiuva nos«. Obsecro, domine, ne desperem suspirando, sed respirem sperando. Obsecro, domine, amaricatum est cor meum sua desolatione, indulca illud tua consolatione. Obsecro, domine, esuriens incepi querere te, ne desinam ieiunus de te. Famelicus accessi, ne recedam impastus. Pauper veni ad divitem, miser ad misericordem; ne redeam vacuus et contemptus
conclusion in M 67 that because it is the image of the Trinity, the soul need only reflect on itself to grow in understanding of God:23

O Lord, I acknowledge and give thanks that You created in me Your image so that I may love, understand, and be conscious of You. But [this image] has been so effaced by the abrasion of vices, so hidden from sight by the dark billows of sins, that unless You renew and refashion it, it cannot do what it was created to do.

The confident rhetorical tone at the end of the Monologion, bordering on insouciance, could hardly have satisfied Anselm’s religious sense, especially the sense of his own unworthiness of salvation as evidenced by his Oratio to Paul. The mood here at the beginning of the Proslogion is more to his liking, but the consolation he hopes to feel forces him to attribute mercy to God, leaving him with the need to explain how the impassive God of his philosophy can be the same as the anthropomorphic God of his prayers. This is the question he examines in chapter 8.

We’ve already looked at Anselm’s reaction to the philosophical problem posed in Isidore’s Etymologies. God cannot be emotionally moved by the suffering of another, since he is immovable. But it is permissible to say that God’s actions can appear merciful to us, even though he in fact does not feel emotion, that God acts as he intended to act from all eternity—saving the wretched and pardoning sinners—while it seems to the saved and pardoned that their cries have softened his heart. This explains why we might call God ‘merciful’, but it doesn’t explain why God acts in this way.

23 P 1 (H&R): Fateor, domine, et gratias ago, quia creasti in me hanc imaginem tuam, ut tui memor te cogitem, te amem. Sed sic est abolita attritioe vitiorum, sic est offuscata fumo peccatorum, ut non possit facere ad quod facta est, nisi tu renoves et reformes eam
It is especially tricky, given the considerations at the end of the Monologion, to discern why Supreme Justice might spare those who merit eternal damnation. Justice seems to require that the good are rewarded with good, and the bad with bad. So how can God not be just? Anselm tries to solve the problem first by proposing a distinction between God’s justice and goodness. As a sinner himself, he believes God’s merciful action to be very good, and so he proposes that God’s goodness is greater if He saves at least some of the wicked than it would be were he to save none. This solution is untenable, though, because God’s goodness cannot be opposed to justice, so His mercy must flow from His justice. Anselm declares that this is “hard to understand,” but “necessary to believe.”

Characteristically, when faced with such puzzles, he pushes further. God’s goodness must be such that nothing can be understood to be better, and his power must be such that nothing can be thought to be more powerful. These premises lead Anselm to propose three arguments for the concordance of mercy and justice in God: (i) There are two ways of being good, the way of retribution and the way of forbearance, and God

24 Eugene Fairweather argues based on this passage that for Anselm, mercy and goodness flow from God’s justice. As I hope I have sufficiently established, the idea that the divine attributes are derived from one another is a provisional way of speaking because they are not consistent with Anselm’s views of divine simplicity and of the identity of the divine attributes. In any case, Anselm will revise this statement in P 10-11, so it is not even the settled position of the Proslogion. Cf. “‘Iustitia Dei’ as the ‘Ratio’ of the Incarnation,” in Spicilegium Beccense I, 1959, pp. 327-335, esp. 331-2.

25 P 9: Nam etsi difficile sit intelligere, quomodo misericordia tua non absit a tua iustitia, necessarium tamen est credere

26 P 9: “How could it be the case [that You are Supremely Just] if you were good with respect to retribution and not forbearance?” Hoc utique non fieret, si esses bonus tantum retribuendo et non parando...
could be thought to be greater if he could be good only through retribution; (ii)
Furthermore, God could be thought to be more powerful if He were unable to bring good
from evil, so he must be able to do so. (iii) Finally, it seems absurd to think that God
would be doing something evil, but since that which is opposed to justice is evil, if saving
the wicked were opposed to justice, then God would be committing evil.

God’s mercy is not in conflict with His justice, but neither does His mercy
necessitate that he spare all the wicked. Anselm thinks that he can get some grasp on why
God saves some but not all:27

For when You punish those who are evil, it is just because it
besuits their merits. But when You spare them, it is just, not
because [sparing them] besuits their merits but because it befits
Your goodness.

With reference to God, it can be fitting, and therefore just, to spare the wicked; the
wicked, however, do not deserve to be spared. So when God spares the wicked, He seems
to men to be merciful though unjust; from His vantage, He is acting not from the emotion
of mercy but simply as befits His supreme goodness. The conflict of justice and mercy in
this case is mostly verbal, and their reconciliation is accomplished by shifting the frame
of reference. Yet, after this verbal “solution” we can still ask why God saves some of the
wicked and not all of them. Anselm throws up his hands at this point—clearly the
Supreme Justice Himself is acting justly, but we cannot comprehend such justice. It is
clear, though, that justice must be explained in terms of the divine will: “That alone is

27 P 10 (H&R, modified): Cum enim punis malos, iustum est, quia illorum meritis
convenit; cum vero parcis malis, iustum est, non quia illorum meritis, sed quia bonitati
tuae condecens est.
just which you will, and is unjust which you do not will."^{28} Thus God is always just to forgive the wicked, because whatever He does must be just, even if his reasons are incomprehensible to us.

The introduction of mercy into chapter 1 throws a wrench into the *Monologion*’s conclusions about divine justice. The confident claim of M 71,^{29}

> Nothing follows more evidently or ought to be believed more certainly than that man’s soul is created such that it will suffer eternal misery if it disdains to love the supreme essence.

yields to the quite different claim of P 11:^{30}

> Surely we cannot at all comprehend the reason why from among those who are of similar wickedness You save some and not others because of Your supreme goodness, and condemn some and not others because of Your supreme justice.

The difference between these two statements is that the second one has to take account of the mercy of God. The view of the *Proslogion* is thus more complicated, and the conclusions are less satisfying. If man cannot know at all why some sinners are damned and others spared, and if, thanks to original sin, all are in some way sinners, then man cannot understand at all (“*nulla ratione*”) what might be his most fundamental relationship with God! Of course, Christian faith teaches that those who identify

---

^{28} P 11: Nam id solum iustum est quod vis, et non iustum quod non vis.

^{29} M 71: Nihil igitur videri potest consequentius et nihil credi debet certius, quam hominis animam sic esse factam, ut si contemnat amare summam essentiam, æternam patiatur miseriam

^{30} P 11: Illud certe nulla ratione comprehendi potest, cur de similibus malis hos magis salves quam illos per summam bonitatem, et illos magis damnes quam istos per summam iustitiam.
themselves with Christ will be saved, but one needn’t know much about Anselm to know that he wants to understand what he at first accepts on faith.\(^3\) As we will see below, the questions left over from this portion of the _Proslogion_ will remain with him for years, structuring his intellectual development to a large degree until he writes his masterpiece, the _Cur Deus Homo_.

6.5 Justice in the ‘Three Treatises’

Anselm published the three dialogues _De Veritate_, _De Libertate Arbitrii_, and _De Casu Diaboli_ together in about 1086, after he was made Abbot of Bec. Anselm describes them as “pertaining to the study of sacred scripture,” but this is somewhat misleading. Although they do each take off from a passage in scripture, the passages only provide the framework of the problem that the dialogue tries to solve. Thus, the opening line of _De Veritate_ contrasts the claim in John 14:6 that Jesus is truth with the common understanding that truth is in many things, and asks whether we are speaking of God

\(^3\) As George Heyer aptly writes, “This attempt in the _Proslogion_ to harmonize divine justice and mercy suffers from certain weaknesses…The result Anselm achieves falls short of the goal, for he has juxtaposed justice and mercy, not harmonized them. He has set the two perfections side-by-side, but their relationship is hardly more than tangential. Certainly he has not shown how God’s mercy rises and flows from the source of his justice. Instead, Anselm has posited two senses of divine justice—one with a view to man’s deserts and one with a view to God’s nature. What are we to think of them? Surely he does not mean that two justices reside in a God of utter self-consistency…. Anselm has driven a wedge between God’s being and his action. It appears in the treatment of compassion, for there, as we noted, God’s acts of mercy make him appear compassionate when really he is nothing of the sort. These acts not only fail to reveal God as he is in himself but are even likely to deceive men about his true nature…to conclude that God feels emotions.” George S. Heyer, Jr., “St. Anselm on the Harmony between God’s Mercy and God’s Justice” in _The Heritage of Christian Thought: Essays in Honor of Robert Lowry Calhoun_ edited by Robert E. Cushman and Egil Grislis”, Harper & Row, 1965, pp. 31-40. The quotation is from pp. 34-5.
whenever we say something is true. Anselm answers the question by appealing to logic and metaphysics in order to arrive at a definition of truth, but in these dialogues he keeps to the same method of proceeding without appeals to authorities that he employs in the *Monologion*. Since Anselm himself is somewhat coy about his purpose in writing these dialogues, I’d like to propose that Anselm is developing the tools for figuring out the problem of justice and its relation to mercy that left him unsatisfied at the end of P 11. I cannot defend this claim in depth here, but I can sketch the defense one might give.

In his preface to the dialogues Anselm declares that he wants them to be arranged in their traditional order, which suggests that there is a progression in argument from *De Veritate* to *De Casu Diaboli*. The topic of justice and the will’s ability to preserve it is the only one that is prominent in all three dialogues, and Anselm’s treatment grows more sophisticated and builds upon the conclusions of the earlier works. In *De Veritate* Anselm defines justice as “rectitude of will that is preserved for its own sake,” and in *De Libertate Arbitrii* he defines freedom as the power to preserve justice understood in this sense. Anselm argues that (creaturely) justice is the sort of thing that cannot be acquired but can only be given. Since God gives rectitude of will so that the creature may keep it, when the creature doesn’t keep it, he loses it (there are not degrees of justice so defined).

In *De Casu Diaboli*, he argues that there are two powers of the uncorrupted will, one that preserves justice and the other that seeks what is advantageous, and that the

---

32 DV 13.

33 DLA 3.

34 DCD 12-13.
decision to sin is the decision to seek one’s advantage without desiring to preserve justice—to seek what seems good to one without regard to God’s right ordering of the cosmos.\textsuperscript{35}

The view of justice defined in DV 12 can thus be said to dominate the later two dialogues. Anselm defines justice the way he does after he develops a metaphysical idea of truth, defined as “rectitude perceptible to mind alone,” which he thinks is a pure perfection and a divine attribute. Anselm clearly thinks rectitude is a pure perfection, and introduces it into the definitions of justice and truth as a move towards finding a reflective equilibrium among the divine names.\textsuperscript{36} In the Cur Deus Homo, we will see, rectitude enters into the description of mercy. Rectitude thus becomes the key to unifying several divine names, ultimately allowing Anselm to explain how God’s justice and mercy are complementary, and to answer the question why God became man. So rather than thinking that these dialogues are about three different topics—the perennial philosophical questions of truth and free will, and the theological topic of the fall of Satan—it is more helpful to see them as an extended meditation on the nature of rectitude in creation, which is nothing other than God’s proper ordering of all things. If this thesis is correct, then it makes sense for us to trace what Anselm has to say about rectitude, beginning with De Veritate.

\textsuperscript{35} DCD 16, 27.

\textsuperscript{36} DV 12: “Truth and rectitude and justice are defined in terms of each other.” “...invicem sese definiunt veritas et rectitudo et iustitia.”
6.6 Truth, Rectitude, and Justice in *De Veritate*

Anselm identifies truth with rectitude in DV 2.37

Teacher: What is an affirmation designed to do?
Student: To signify that what-is is.
T. Then, this is what an affirmation ought to do…
T. So when an affirmative statement signifies that what-is is, it signifies what it ought to…
T. But when it signifies what it ought to, it signifies rightly, or correctly…
T. And when it signifies correctly, its signification is correct…
T. Therefore, when it signifies that what-is is, its signification is correct…
T. Moreover, when it signifies that what-is is, its signification is true…
S. Yes, its signification is both correct and true when it signifies that what-is is.
T. So for an affirmation to be correct is the same as for it to be true, namely, for it to signify that what-is is…
T. Therefore, the affirmation's truth is simply its rightness, or correctness (*rectitudo*).
S. I now see clearly that truth is this rightness.

Truth in statements is when the statements signify correctly what they ought to signify.

Rectitude here is a teleological notion: “when it signifies what it ought to, it signifies correctly (*recte*)” argues the Teacher persuasively. In DV 4, the Student concludes that

rectitude in the will, defined as when one wills what one ought, is also teleological. Likewise, rectitude in action is doing what one ought or even suffering what one ought or being what and where one ought. Rectitude can be either natural or non-natural, depending on whether a thing’s telos is determined by the nature of the thing or something outside the thing. So a true statement like “It is day” has two kinds of truth, a natural truth if it is a well-formed sentence (it fulfills the nature of a sentence, i.e., to signify), and a non-natural truth if it signifies that what-is is so. Likewise, rectitude can be either contingent or necessary, depending on whether the rectitude depends on the will or not: Anselm says that a fire’s heating is necessary, but a man’s doing good is not from necessity.

---


39 DV 5: Sed sententia est omnium quia qui facit quod debet, bene facit et rectitudinem facit.

40 Ibid. (H&R): “However, when the Lord said, “He who does the truth comes to the light,” He wanted us to understand the verb “to do” not only as standing for what is properly called a doing, but also as a substitute for every other verb. For He excludes from this truth, or light, neither the man who undergoes persecution for the sake of justice nor the man who is when and where he ought to be, nor the man who is standing or sitting when he ought to—and the like. For no one denies that such persons do what is good.” “Facere autem non solum pro eo quod proprie dicitur facere, sed pro omni verbo dominus voluit intelligere, cum dixit quoniam »qui facit veritatem, venit ad lucem«. Non enim separat illum ab hac veritate sive luce, qui patitur persecutionem »propter iustitiam«; aut qui est quando et ubi debet esse; aut qui stat vel sedet quando debet; et similia. Nullus namque dicit tales non bene facere.”

41 DV 5.

42 Even though, as Anselm argued in M 68, rational nature was created for loving God in the same teleological sense that fire received its ability to heat. It would seem on
To this point, Anselm has been telling us how to use the concept of rectitude, but he still needs to explain its fundamental teleology. He takes up this question in chapters 7-8. In a passage reminiscent of Boethius’s *De Hebdomadibus*, Anselm argues that truth is in all things insofar as they have received their being from God. Insofar as a thing corresponds to that purpose for which God designed it, we say it is true. Since a thing ought to be as God made it to be, and a thing is more fully a being if it corresponds to its paradigm in the Supreme Essence, a thing is rightly insofar as it is. Thus, if truth and rightness are in the being of things because these things are what they are in the Supreme Truth, then assuredly the truth of things is rightness (rectitudo).

As Marilyn Adams argues, this sense of truth being “in” all things and things being “in” the Supreme Truth hearkens back to the analogy in M 10 between God and the artist who creates according to the image he has in his mind. For Anselm, God is the source of the truth in a thing, and thus the standard by which its rectitude can be judged.

But, the Student asks, if all things get their being from God, and some things do Anselm’s view that man’s being good is contingent but natural. Marilyn McCord Adams, in her otherwise excellent article on *De Veritate*, claims that natural and necessary rectitude are identical. See “Saint Anselm’s Theory of Truth,” p. 363.

43 DV 7: M. An putas aliquid esse aliquando aut alicubi quod non sit in summa veritate, et quod inde non acceperit quod est inquantum est, aut quod possit aliud esse quam quod ibi est? D. Non est putandum.

Adams (“Truth”, p. 364) points out that this passage recalls the Student’s statement in ch. 2 that “nothing is true except by participating in the truth. Therefore the truth of something true is in the true thing itself.” This seems right. What’s more, if this passage is an answer to the question of ch. 2, the mention of “participation” adds a textual clue to the suggestion from doctrinal similarity that this passage, as with M 16, is influenced by Boethius’ *De Hebdomadibus*.

44 Ibid. (H&R): M. Si ergo et veritas et rectitudo idcirco sunt in rerum essentia, quia hoc sunt quod sunt in summa veritate: certum est veritatem rerum esse rectitudinem.
not act as they ought, then it would seem that God is the source of their injustice. The Teacher responds that God cannot be the source of injustice; indeed, everything that happens God causes (or permits to happen) wisely and well, and everything that comes from such Goodness certainly ought to be.\textsuperscript{45} This is the typical medieval version of the classical problem of evil,\textsuperscript{46} which Anselm attacks by deploying the two senses of rectitude he developed earlier.\textsuperscript{47} It is true both that Christ, as supremely innocent, ought not to be killed and that no one ought to have killed him. Let’s call this \textit{rectitude for the subject}, which is a species of natural rectitude. It is also true that he ought to have been killed, because he (who is God) wished to suffer it, and for any state of affairs \(x\), if God wishes \(x\), then \(x\) ought to be. Let’s call this \textit{global rectitude}.\textsuperscript{48} So Jesus both ought and ought not to have been killed, depending on the point of view.\textsuperscript{49}

\textsuperscript{45} DV 8.

\textsuperscript{46} As opposed to the contemporary problem of evil. The classical problem was to explain the metaphysics of evil, given the goodness of God and the convertibility of being and goodness. The contemporary problem is to explain the existence of a benevolent and providential God given the existence of evil.

\textsuperscript{47} In what follows, I follow very closely upon John O’Neill’s “‘The Same Thing Therefore Ought to Be and Ought Not to Be’: Anselm on Conflicting Oughts,” in \textit{Heythrop Journal} 35 (1994), pp. 312-314.

\textsuperscript{48} As we saw earlier, Anselm is also committed to the view that God wills every actual state of affairs, and therefore any actual state of affairs is just and (globally) right.

\textsuperscript{49} This solution is reminiscent of the solution to the problem in P 9-11 of whether God is just in saving the wicked. In that passage we saw how Anselm proposed merely a verbal answer to the seeming injustice of God’s pardoning the wicked, by defining two senses of justice, one to explain God’s point of view and one to explain man’s. The distinction here, however, is not merely verbal, because it gives a unified account of what God wills (the state of affairs and all the natural causes of the state of affairs) and what he merely permits (that rational creatures use their natural powers freely, even if they don’t act according to their natural ends).
This distinction allows Anselm to explain the problem of evil: God always wills global rectitude and grants creatures natural rectitude, while sin is the voluntary abandonment by a rational creature of his natural rectitude. Just as the nails that were driven into the flesh of Christ were good nails, and thus according to their nature ought to penetrate the flesh, so too is Pilate’s will to act unjustly a good will, because according to its nature it is capable of deciding freely. Just as the innocent Christ ought not to be crucified because it is not fulfilling for men to be crucified and it is not just that innocent men should be treated as criminals, so Pilate ought not to sin in crucifying Christ because he was created to will justice. Finally, just as God’s willing that the nails behave as nails and that the innocent Christ be crucified is just, because everything God does is just, so is it just for God to will that Pilate be able to crucify Christ and that he actually do so. It is not that Pilate doesn’t sin—he does. But Pilate’s sinning is part of God’s plan, and therefore of global rectitude, even though it is no part of Pilate’s rectitude, since it was not right that his sin be the cause of an innocent man being put to death.

The idea of global rectitude will lead Anselm to reflect on God’s foreknowledge and providence, the topic of De Libertate Arbitrii. But before we turn there, we should examine Anselm’s definition of justice in DV 12. The Student is dissatisfied with the simple identification of justice and rectitude. Although such identification surely makes sense for God,\textsuperscript{50} it would seem to mean that even a stone could be just if it behaves like a proper stone. The Teacher then adds to the definition of justice that the rectitude must be

\textsuperscript{50} DV 12: “In the highest and simple nature, there is no doubt that justice and rectitude are the same.” “In summa namque et simplici natura…dubium tamen non est idem esse rectitudinem et iustitiam.”
praiseworthy, which implies that “for men” it is found in the will, and that not just the object willed but also the manner of willing must be correct. One can will to do what one ought, but out of a desire for vainglory or fear of punishment, rather than because it is the right thing to do. Thus, the Teacher concludes, justice is rectitude of will preserved for its own sake. In keeping with his procedure of reflective equilibrium, he tests to see whether this version of justice is consistent with the divine nature, and is thus a pure perfection.  

T. Do you think that this definition can be applied to the Supreme Justice—insofar, that is, as we are able to speak about a thing of which nothing, or almost nothing, can properly be said? S. Since in God's divinity, power is not other than the divinity itself, we speak of the power of His divinity or of His divine power or of His powerful divinity. Similarly, although God's will is not one thing and His uprightness another, nevertheless it is not unsuitable for us to speak of His uprightness of will or of His voluntary uprightness or of His right will. But if we say that God's uprightness is kept for its own sake, then we seem not to be able to say this as suitably about anyone else's uprightness. For just as nothing else keeps God's uprightness, but it keeps itself, and just as it keeps itself through nothing other than through itself, so it keeps itself for the sake of nothing but itself. T. Then, we can say with certainty that justice is uprightness-of-will which is being kept for its own sake.

---

51 Ibid. “in homine”

52 Ibid. (H&R): M. Videtur tibi quod ista definitio possit aptari summæ iustitiae, secundum quod de re loqui possimus de qua nihil aut vix aliquid proprie potest dici? D. Licet non ibi sit aliud voluntas, aliud rectitudo, tamen sicut dicimus potestatem divinitatis aut divinam potestatem sive potentem divinitatem, cum in divinitate non sit aliud potestas quam divinitas: ita non inconvenienter dicimus ibi rectitudinem voluntatis aut voluntariam rectitudinem seu rectam voluntatem. Si vero illam rectitudinem dicimus propter se servari, de nulla alia rectitudine sic convenienter dici posse videtur. Sicut enim non aliud illam sed ipsa se servat, nec per aliud sed per se: ita non propter aliud quam propter se. M. Indubitanter igitur possimus dicere quia iustitia est rectitudo voluntatis, quæ rectitudo propter se servatur.
Two points can be made here. 1) Anselm here fixes the mistake he introduced in P 11, where he (inexplicably) distinguished between the nature and the activity of God.\(^5\) Now, he argues, since God’s will is the same as His rectitude, which is the same as His nature, there is no question that God wills in accord with His nature as He ought. Although will and nature and rectitude are not distinct in God, there is no harm in speaking about them separately, and so the definition of justice for men can also apply to God, *mutatis mutandis*. 2) We should not discard the sense of rectitude and justice that applies to the stone. If we let justice in a thing be defined as ‘something doing everything that it ought to do according to its nature’, then we have one definition of justice for the stone, for man, and for God. Because rational natures are more complex than non-rational natures, the demands of justice are greater for men than for stones—they have to have rightness of will for its own sake, which has its ultimate basis in the rectitude of God. But as we saw in M 68, man was created for justice. Therefore, for man to will justice is simply for him to do that for which he was created, just as a stone behaves with rectitude when it does that for which it was created. Although later scholastics would not put it quite this way, it seems that Anselm is aiming for a definition of justice as a genus that can be divided into three species: justice for the non-rational, justice for rational creatures, and justice for God.\(^5\)

---

\(^5\) Cf. Heyer’s observation on P 11, quoted above: “Anselm has driven a wedge between God’s being and his action.”

\(^5\) In the next chapter, Anselm even seems to say that justice is a genus: “Let us return to rectitude or truth, two names by which (since we speak of [truth as] rectitude perceptible only to the mind) one thing is signified which is the genus of justice.” “M. Redeamus ad rectitudinem seu veritatem, quibus duobus nominibus, quoniam de rectitudine mente sola perceptibili loquimur, una res significatur quae genus est
At the beginning of *De Libertate Arbitrii*, the Teacher insists that ‘freedom’ must be said equally of man, the good angels, and God, after the fashion of the genus ‘animal’:  

Student. Since free choice seems to be opposed to the grace, predestination, and foreknowledge of God, I desire to know what

---

freedom of choice is and whether we always have it. For if freedom of choice consists in being able to sin and not to sin (as some persons are accustomed to say) and if we always have this ability, how is it that we sometimes need grace? But if we do not always have this ability, why is sin imputed to us when we sin without a free choice?

Teacher. I do not think that freedom of choice is the ability to sin and not to sin. Indeed, if this were its definition, then neither God nor the angels who are not able to sin would have free choice—a blasphemous thing to say.

S. What if we say that the free choice of God and of the good angels is different from ours?

T. Although the free choice of men differs from that of God and of the good angels, nevertheless the definition of this freedom ought to be the same in both cases, in accordance with the name “freedom.” For example, although one animal differs from another either substantially or accidentally, the definition [of “animal”] is the same for all animals, in accordance with the name “animal.” Hence, it is necessary to give such a definition of “freedom of choice”—a definition which contains neither more nor less than freedom does. Therefore, since the free choice of God and of the good angels is not able to sin, “to be able to sin” does not pertain to the definition of “freedom of choice.” In fact, the ability to sin does not constitute either freedom or a part of freedom. To understand this point clearly, pay attention to what I am going to say.

Since God and the good angels cannot sin, freedom must not be the ability to sin or not to sin. Clearly, Anselm intends in this dialogue to follow the same procedure he followed in De Veritate—first he will examine everything that is said to be ‘free’, and then he will develop a definition that will apply equally to all free things, and this will be the definition of the genus ‘free things.’ That he intends to treat ‘libertas’ as a genus is clear from the last two chapters of DLA, in which the Teacher is clearly trying to provide a classical Aristotelian division of a genus into its species.⁵⁶

---

Anselm clearly has an agenda that is incompatible with the Student’s seemingly sensible proposal that they treat freedom as an analogous term, with angelic and divine freedom being in some way different from human freedom. Why is he so set on finding a single definition of freedom, when doing so seems opposed to the common usage? This seemingly eccentric requirement starts to make sense when we consider the single definition of justice in DV 12. The argument in that chapter commits Anselm to the following premises:

L1: Justice for rational beings is rectitude of will preserved for its own sake.
L2: Justice for rational beings must be worthy of praise.
L3: Free choice of the will alone makes preserving rectitude worthy of praise.

From which it follows,

1. Justice for rational beings requires free choice of the will.
2. Insofar as the definitions of freedom among the rational beings are different, to that degree the definitions of justice will be different as well.

But if there are radically diverse definitions of justice for God and man, then the whole attempt at understanding God’s nature through understanding the pure perfections breaks down. So Anselm is committed by the whole thrust of his rational theology to the denial of the Student’s suggestion.

Instead, Anselm argues that freedom is always some capacity, and freedom of will is the capacity to do what the will ought to do, which is to preserve rectitude for its own sake. This capacity is part of human nature, even in sinners, who lack rectitude of will—while they lack both rectitude and the ability to acquire it on their own, they always

57 DLA 3: “omnis libertas est potestas”
have the power to keep it if it is restored to them.\textsuperscript{58} This power is such that nothing can overcome it—as we said in an earlier chapter, a man can be bound unwillingly, he can be killed unwillingly, but he cannot will unwillingly, and so he is always responsible for what and how he wills. Thus, free will cannot be overcome by another power,\textsuperscript{59} even by strong temptations,\textsuperscript{60} or even God.\textsuperscript{61} A will is said to be a “slave to sin,”\textsuperscript{62} because it cannot acquire rectitude on its own, but since it never loses the power by which it could preserve rectitude were it to reacquire it, it is still free. Since this freedom is a power (potestas), it can appropriately be said even of God—or especially of God, who is all-powerful.\textsuperscript{63}

Since much of this dialogue consists in tying up loose ends of the definition of justice from \textit{De Veritate}, it needn’t detain us further. Anselm develops his ideas about the power of the will at length in \textit{De Casu Diaboli} and in \textit{De Concordia}, and there is some evidence that his theories of the will were considered to be particularly innovative in his day.\textsuperscript{64} We don’t need an in-depth investigation for our purposes, but we do need to look

\textsuperscript{58} DLA 4.

\textsuperscript{59} DLA 5.

\textsuperscript{60} DLA 6-7.

\textsuperscript{61} DLA 8.

\textsuperscript{62} John 8:34.

\textsuperscript{63} Cf. DLA 14.

\textsuperscript{64} In his autobiography, Guibert of Nogent (1055-1125) tells of several meetings between the young Guibert and Anselm when he was prior of Bec. Guibert writes that Anselm’s teaching “was to divide the mind three- or fourfold, to treat the operations of the entire inner mystery under the headings of appetite, will, reason, and intellect. He
at a few features of *De Casu Diaboli*, justice and rectitude start to acquire a little more content.

6.8 Freedom and the Right Ordering of Creation: *De Casu Diaboli*

At this stage Anselm’s description of justice is too formal for it to be satisfying. We can see this if we try to arrange our definitions of justice from their most abstract to the most concrete:

1. Justice is rectitude.
2. Justice is doing what God wants done in the manner he wants it done.
3. Justice is a thing conforming to its nature, which either is God or is from God.
4. Justice (for rational beings) is rectitude of will preserved for its own sake.

There can probably be disagreements about the ordering of 2-4, but it is clear that all of them are fairly abstract definitions. That they are so abstract helps Anselm derive necessary conclusions from them—while it would seem impossible to prove that, for example, the duty of Sabbath observance must follow from the nature of God, it seems possible, at least, to prove that the duty to pursue justice might so follow. Here is a typical exchange, where the Teacher is trying to explain how the devil first sinned:

showed that the first two—appetite and will—which most people, including myself, consider to be one…are not in fact identical, even though one can readily assert that from the viewpoint of the intellect they are practically the same.” [*A Monk’s Confession: The Memoirs of Guibert of Nogent*, transl. Paul J. Archambault (Pennsylvania State University Press, 1996) p. 61.]

65 *DCD 4* (H&R, with small changes): M. Quidquid habebat, debebat velle. D. Vere debebat velle quod a deo acceperat, nec hoc volendo peccavit. M. Voluit igitur aliquid quod non habebat nec tunc velle debebat…M. Nihil autem velle poterat nisi iustitiam aut commodum. Ex commodis enim constat beatitudo, quam vult omnis rationalis natura. D. In nobis hoc possimus cognoscere, qui nihil volumus nisi quod iustum aut commodum putamus. M. Iustitiam vero volendo peccare non potuit… M.
T. Whatever he had, he was supposed to will.
S. Yes, he was supposed to will what he had received from God, and he did not sin by willing that.
T. Therefore, he willed something which he did not already have and was not supposed to will at that time...
T. But [the Devil] was able to will nothing except what is just or beneficial. For happiness, which every rational nature wills, consists of benefits.
S. We can recognize this in ourselves, for we will nothing except that which we think to be just or beneficial.
T. But [the Devil] was not able to sin by willing justice...
T. Therefore, he sinned by willing something beneficial which he did not possess and was not supposed to will at that time but which was able to increase his happiness...
T. You recognize, I think, that by inordinately willing more than he had received, he extended his will beyond justice …
T. But when [the Devil] willed what God did not will him to will, he willed inordinately to be like God.
T. Even if he did not will to be altogether equal to God, but contrary to the will of God willed to be something less than God, then even in this case he willed inordinately to be like God; for he willed something by his own will (propria voluntate), which was subject to no one else. For it ought to be the characteristic only of God so to will something by his own will such that He is not subordinate to a higher will…
T. However, not only did [the Devil] will to be equal to God because he presumed to have an autonomous will, but he even willed to be greater [than God] by willing what God did not will him to will, for he placed his will above the will of God.

Since all creatures have been created for some end, and that end is given by God, to desire more than has been given by the supremely good and wise God is to sin, and in

Peccavit ergo volendo aliquod commodum, quod nec habebat nec tunc velle debuit, quod tamen ad augmentum illi beatitudinis esse poterat… M. Cernis, ut puto, quia plus aliquid quam acceperat inordinate volendo voluntatem suam extra iustitiam extendit… M. At cum hoc voluit quod Deus illum velle nollebat, voluit inordinate similis esse deo… M. Etiamsi noluit omnino esse par Dei, sed aliquid minus deo contra voluntatem dei: hoc ipso voluit esse inordinate similis deo, quia propria voluntate, quæ nulli subdita fuit, voluit aliquid. Solius enim dei esse debet sic voluntate propria velle aliquid, ut superiorem non sequatur voluntatem… M. Non solum autem voluit esse aequalis deo quia præsumpsit habere propriam voluntatem, sed etiam maior voluit esse volendo quod Deus illum velle nollebat, quoniam voluntatem suam supra dei voluntatem posuit.
two ways—materially, by willing what ought not to be willed, and formally, by not subjecting the created will to the just will of God. Anselm shows that when the devil was not satisfied with his God-given abilities and place in the universe, it was because he conceived of and desired some better lot than he had (being an angel, he had no passions to lead him to lower pleasures). In doing so he not only refused to submit to God, he in effect tried to usurp God’s role as the source of justice, or more precisely, he committed the sin of *presumption*—the act of asserting rights to a place in an ordered society to which one is not entitled. This is a nifty piece of reasoning, and it introduces Anselm’s famous “two powers of the will doctrine,” but it doesn’t tell us what Satan wanted. Was he a snob, upset to learn that men were to join the angels in heaven? Did he just resent playing second fiddle to God? To be sure, it isn’t Anselm’s purpose to write *Paradise Lost*. But it is his purpose to learn something about God’s justice, and such a purely formal idea of justice borders on the inscrutable.

For the larger purposes of this dissertation, a formal idea of justice is all that is necessary. But Anselm would like something more. He gets a little bit more in chapter 16, where argues that when God grants justice to a rational being, he gives it a great dignity that it ought to keep. If that rational being loses its dignity, it is worse than it would have been if it had never had it, for it is lacking what it ought to have.$^6$

---

$^6$ DCD 16 (H&R): “D. Quoniam nihil aliud est additum quam iustitia, separat a iustitia nihil aliud certum est remanere quam quod prius erat, nisi quia debitricem eam fecit accepta iustitia, et quasi quaedam pulchra vestigia sui reliquit derelicta eadem iustitia. Eo enim ipso quo debitrix permanet iustitiae, monstratur honestate iustitiae fuisset decorata. Sed et hoc satis iustum est ut quod semel accepit iustitiam, semper eam debet…”
Since only justice was added, then when justice is gone, surely there remains only what was first there—except for the fact that having received justice has made the will a debtor, and, so to speak, certain lovely vestiges of justice have remained after justice has been deserted. For by the very fact that [the will] remains a debtor to justice, it is shown to have been adorned with the honor of justice. But even this is sufficiently just, viz., that what once received justice should always be under obligation with respect to justice…

When the rational being has its dignity, therefore, it owes a debt to God and to itself. If it loses its dignity, it still has the debt—it still ought to have its proper dignity—but now it is diminished, possessing only a ‘vestige’ of the dignity proper to it. A ‘vestige’ is a little more than mere ‘privation.’ If even fallen creatures have a vestige of justice, a remnant of the dignity that once adorned their nature, perhaps this might allow us to draw back the veil of abstraction surrounding the notion of justice. If there is a hint of rectitude even in those who have lost it, then even in a creature that can only pursue its own happiness, there is at least a sense of something more, something missing. This might explain the sense of longing for God that Anselm describes at the beginning of the *Proslogion*, which in P 17 leads him to posit God’s sweetness, harmony, and general desirability.

If we recall the doctrine of DV 8, the rational being ought to have dignity in two sense of ‘ought’: 1) because God in his providence gave it, and everything God wills ought to be; and 2) because it is natural that man be so dignified, and man ought to fulfill his nature. So the creature that loses that dignity is blameworthy twice-over.67

---

67 DCD 16 (H&R): D. Habuisse vel debere monstrat naturalem dignitatem, et non habere facit personalem inhonestatem. Debere enim factum est ab eo qui dedit, non habere vero factum est ab ipso qui deseruit. Ideo namque debet quia accepit, ideo vero non habet, quia dereliquit. M. Non ergo reprehendis in ipsa voluntate que non stetit in iustitia debere iustitiam, sed non habere iustitiam. D. Omnino nihil aliud ibi reprehendo
S. To have had, or to be under obligation with respect to, [justice] manifests a natural dignity; not to have [justice] causes personal dishonor. For owing was caused by the one who gave; but not having was caused by the one who deserted. For he is indebted because he has received; but he does not have because he has deserted.

T. Therefore, what you are blaming in the will which did not stand fast in justice is not that it owes justice but that it does not have justice.

S. What I blame in that will is nothing at all other than the absence of justice, or the not-having of justice. For, as I have already said, to owe [justice] beautifies; but not to have [justice] mars. And the more the debt is becoming, the more the not-having is unseemly. Indeed, the will is marred by the not-having, for which it itself is to blame, only because it is adorned by the obligation-to-have, which results from the goodness of the Giver.

Anselm is very alive to the deep connection between ‘ought’ and ‘owes’68; and so he characterizes the obligation to fulfill one’s God-given nature as a ‘debt’ to God.69 Being under a debt of this sort is not a bad thing; indeed, being under a debt to God is a good thing, as is everything willed by God. Only if a rational creature lacks what it ought to have is the debt burdensome. He has no trouble deriving an ‘ought’ from an ‘is’ here

quam absentiam iustitiae, sive non habere iustitiam. Nam sicut iam dixi debere decorat, non habere vero deturpat; et quo magis illud decet, eo magis istud dedecet; immo non ob aliud illam voluntatem deturpat ex propria culpa non habere, quam quia illam decorat ex bonitate dantis debere habere.

68 In Latin, the word ‘debere’ means both ‘ought’ and ‘owes.’

69 John McIntyre makes this point very nicely: “It must also be pointed out that for St. Anselm the notions of ‘ought’ and ‘owing a debt’ are the same, so that the propositions: ‘the rational nature ought to subject its will to the Will of God,’ and ‘the rational nature owes God the debt of subjecting its will to the Will of God,’ are logically equivalent…. That is, whereas we nowadays might seek other grounds for our obedience to God, suggesting that the debtor-creditor category is inadequate to describe the relationship of creature to Creator, St. Anselm finds the sole reason for such obedience in the fact of man’s owing God that debt, or, more accurately still, St. Anselm conceives such obedience to be ipso facto the debt that we owe.” St Anselm and His Critics, p. 67.
because the ‘ought’ is a description of an objective state of affairs—that what the devil is lacks something that his nature is capable of and which it is right for him to have.

Anselm’s picture of the sin of Satan is that it arises from a combination of necessary ignorance and desire for illicit advantage. If Satan knew for certain that he would be punished for any insubordination, he could not rationally have believed that to sin would be to his advantage, in which case his decision not to sin would not have been morally praiseworthy. So it is necessary that all the angels not know that God in His Justice would punish any injustice, or else their obedience would not be praiseworthy, but would be out of self-interested advantage. Yet, if the Student can figure out that an injustice was likely to be punished by Justice itself, surely this was evident to the angels.

In response to this worry, the Teacher paints a revealing picture:

But suppose someone should claim: [That angel] was not at all able to believe that God was going to condemn, because guilty, His own creature whom He had created by means of such great goodness. [And he] especially [could not have believed it] because [of the following considerations]: No example of justice punishing injustice had previously occurred. Moreover, he would have been certain that the number of those who had been created to enjoy God was fixed by such great wisdom that just as it had no

Cf. DCD 13.

DCD 23 (H&R): M. … Sed et si quis dicat quia nullatenus credere potuit deum creaturam suam propter eius culpam damnaturum, quam tanta sua bonitate fecerat—præsertim cum nullum exemplum iustitiae ulciscentis iniustitiam præcessisset, et certus esset numerum in quo illi facti erant qui deo frui deberent tanta sapientia esse præstitutum, ut sicut nihil habebat superfluum ita si minueretur imperfectus esset, nec tam præclarum opus dei ex aliqua parte permanesurum imperfectum, nec ulla ratione scire posset si homo iam factus erat, deum humanam naturam pro angelica aut angelicam pro humana si caderet substituturum, sed potius unamquamque in id ad quod facta erat pro se non pro alia restituturum, aut si factus nondum erat homo, multo minus putare posset ad substitutionem alterius naturæ illum esse faciendum—, si inquam hæc aliquis dicat: quæ inconvenientia inest
superfluity, so if it were diminished it would be imperfect; but so excellent a work of God would not remain imperfect in any respect.

Now, if man had already been created, then [that angel] would not at all have been able to know that God was going to substitute human nature for angelic nature, or angelic nature for human nature, if either were to fall. Rather, he would have believed that God was going to restore each nature to that end for which it had been created—restore each to its own place, not to the other's place. On the other hand, if man had not yet been created, then [the angel] would have been all the less able to suppose that man was going to be created as a substitute for angelic nature. Now, if someone should make this claim, what unfittingness would there be in it?

This conjecture assumes that God’s nature is incomprehensible even to angels, because of what we called ontological limits of all creatures. Presumably, the devil was told that and how he should obey God, but perhaps the reason for this command was not fully explained—perhaps it was not explainable due to the ontological limits on creaturely knowledge.

More importantly, this conjecture proposes a surprising understanding of God’s purposes in creation. The devil, Anselm speculates, might have known that God intended that He be worshipped by a particular number of rational natures. Since God knows from all eternity exactly how many rational beings will be worshipping Him, it is not too far-fetched to conclude that in the act of creation He intended this many rational beings to be worshipping Him. Surely, Satan might reason, when God created the angels, He intended for them all to worship Him; therefore, even though the Supreme Justice is able to punish the disobedient, He wouldn’t dare cast them away from Him into Hell, for that would

72 Anselm alludes to Romans 11:33.
thwart His own purposes in creation. “I’m too important to God. He won’t punish me because he needs me,” is the essence of Satan’s argument.

If Satan so reasoned, he was fantastically wrong, it is clear. But we should note that Anselm thinks almost all the premises of the devil’s argument are true. God does create with the intention that a fixed number of rational beings will worship Him. It would thwart His purposes were He to condemn all the beings that disobey Him. The only difference is that the devil cannot imagine God would substitute men for angels who disobey.

This claim that there is a fixed number of rational beings that must worship God, and that any old rational beings will do, is a bit curious. Why must God use the genus ‘rational being’ rather than pick out the particular rational beings that will worship Him? All the rational beings are free, for sure, but God, being eternal, would still know which of them is going to sin. John McIntyre in dismissing this feature of Anselm’s thought, writes, “This argument is in the spirit of the times and it need not delay us further.” It certainly would not persuade many people today, but it is nonetheless significant for its parallels with the doctrine of the Atonement that Anselm ultimately develops. Satan fell because he thought that God’s love for His creatures and His commitment to the plan of His creation bound Him not to punish any angels. In the Cur Deus Homo, Anselm holds that God’s love for His creatures and for the world, and His providential ordering of creation, commits Him to the salvation of at least some men in order that there be enough rational beings in heaven. As McIntyre argues, this is not Anselm’s primary strategy for

73 St. Anselm and His Critics, p. 82.
explaining the Atonement,\textsuperscript{74} nor is it for explaining the fall of the devil—recall that in this passage, the argument is flagged as a tentative conjecture. But it is a strategy to which he returns in the CDH, and it comes from his conviction that God’s providence must extend to every detail of creation and cannot be thwarted. Let us begin our discussion of the \textit{Cur Deus Homo} by looking at how Anselm sets up the analogous problem about whether God in His Justice should punish human beings for their sins.

6.9 Justice as the Limit of Mercy: Honor and Rectitude in the \textit{Cur Deus Homo}

\textquotedblleft \textit{Cur Deus Homo}\textquotedblright means literally \textquotedblleft Why the God-Man.	extquotedblright It is usually translated \textquoteleft Why God became Man," but the real conclusion of Anselm’s treatise is that the salvation of man can only be accomplished by a \textit{deus homo}, a person who unites divine and human natures.\textsuperscript{75} Anselm’s preface to the dialogue explains that he divided it into two books with two different purposes in each book. In the first book, which focuses on God’s strict justice, he endeavors to show those who think that the incarnation is \textquoteleft repugnant to reason” and thus plays into the hands of the infidels, that apart from Christ, no member of the human race could ever be saved. The second book, which focuses on God’s mercy, argues that human nature was created to enjoy blessedness forever, body

\textsuperscript{74} Ibid., p. 81

\textsuperscript{75} Cf. McIntyre, \textit{St. Anselm and His Critics}, p. 117. An alternative explanation of the lapidary title is that it is an abbreviation of the question Anselm begins with, \textquoteleft qua scilicet ratione vel necessitate \textit{deus homo} factus sit?\textquoteright usually rendered \textquoteleft By what reason or necessity did God become man?\textquoteright (The sentence might also be translated, \textquoteleft By what reason or necessity would the God-man be made?\textquoteright, but that would break the link with the Nicene Creed’s \textquoteleft homo factus est,\textquoteright which Anselm’s audience would recognize immediately.) It might be possible that both titles are intended.
and soul, and that it is necessary, given the considerations of the first book, that this be done only through a God-Man.  

The doctrine of the work as a whole requires both books, therefore, since it is Anselm’s purpose to show that God is both just and merciful; most of the criticisms of this heavily criticized work are the result of focusing too much on the first book and neglecting the second.  

Because our purpose here is not to recreate the dramatic despair of the end of the first book (Boso, Anselm’s prize student in real life and interlocutor in the dialogue, exclaims “What appears to be the case from all this is that the mercy of God and the hope of man are dead!”), or the relief that comes from the second, we will proceed more analytically, moving back and forth among the chapters to piece together the relevant parts of Anselm’s full doctrine.

Curiously, Anselm begins the treatise with an appeal to God’s violated honor. This is language unlike any he has used before; usually Anselm talks about rectitude or what is owed to God or what is befitting His dignity. Yet, as we will show in chapter seven, this is a shift in terminology, but not in doctrine—honor and rectitude are practically synonyms for Anselm. The language of ‘honor’ serves to emphasize God’s provident governance, His obligation to order everything rightly, the political side of rectitude.

76 CDH, Praefatio.

77 This one of John McIntyre’s major complaints (cf. St. Anselm and His Critics, especially ch. 2).

78 CDH 1.23

79 CDH 1.11.
Anselm begins I.12, the beginning of the main argument of the treatise, by asking whether it is “fitting” or “decorous” for God to dismiss man’s sins out of mercy alone, without asking for a restitution of the honor taken from Him. Forgiveness here simply means not inflicting punishment, i.e., being lenient, and therefore is analogous to the conjecture from DCD 23 in which Satan assumes he will be not be punished for his disobedience. Anselm argues at length that this sort of leniency is a violation of God’s justice, because it would lead to His kingdom being radically disordered:

A…. Now, in the absence of satisfaction, to order sin rightly (recte) is just to punish it; therefore, if sin is not punished, something disordered is forgiven….
A. But it is not fitting that God should forgive something that is disordered within His kingdom….
A. Therefore, it is not fitting that God should forgive sin that goes thus unpunished.

Anselm’s concern is that God’s kingdom be ordered rightly, as it ought, i.e., according to global rectitude, which requires that as the supremely just ruler, God should not be allowed to lose His honor. Since nothing is greater than God, “then Supreme Justice (which is identical with God Himself) keeps nothing more justly than God's honor in regard to the governance of things.” Given this requirement of justice, were God to

---

80 CDH I.12 (H&R): “A. Redeamus et videamus utrum sola misericordia, sine omni solutione ablati sibi honoris deceat deum peccatum dimittere.

81 Ibid.: “A. Sic dimittere peccatum non est aliud quam non punire.”

82 Ibid.: “A…. Et quoniam recte ordinare peccatum sine satisfactione non est nisi punire: si non punitur, inordinatum dimittitur. A. Deum vero non decet aliquid inordinatum in suo regno dimittere. A. Non ergo decet deum peccatum sic impunitum dimittere.

83 CDH I.13: “Si deo nihil maius aut melius, nihil iustius quam honorem illius servat in rerum dispositione summa iustitia, quae non est aliud quam ipse deus.”
allow His honor to be violated without punishment or satisfaction, it would show that He lacked the power to defend His honor (a silly thought) or that He wasn’t just (but we just said that He is Justice itself).

If we recall the argument of DCD 23, Satan is thought to have believed that God would not punish him because the only way he could be punished is by being removed from God’s presence, and since that would reduce the number of angels in heaven, it would thwart God’s plan for creation. Satan was wrong, it was supposed, because God intended to replace the fallen angels with humans, and so even if Satan is punished, the right number of rational souls in heaven would be achieved. Now, it turns out, the same story repeats itself with men: they have sinned, and God ought not to grant them eternal happiness unless they are punished or can grant Him satisfaction. One difference from DCD 23 is that the fallen angels aren’t offered the option of satisfaction — Anselm only considers whether they are punished or not. Anselm has reasons for this asymmetry between angels and men, which we will examine shortly, but one immediate consideration is that if men are not given the possibility of giving God satisfaction, then God is forced either to elevate to heaven some lower creature on the great chain of being (dolphins, perhaps?), or else not be worshipped properly. But only rational beings can worship God freely, which leaves only men and angels. So it seems that God either punishes man for his sins, thwarting His plan, or is lenient, which also violates rectitude, or else is offered satisfaction by man.

84 This model is repeated in CDH I.16-19.
Anselm reaches the same conclusion with another argument, this time focusing on man’s dignity in isolation from God’s replacement of the angels:

A. Let us suppose that some rich man is holding in his hand a pearl of great value which no impurity has ever touched and which no one else can remove from his hand unless he permits it. And let him be intending to hide it away in his treasury, where his most cherished and most valuable possessions are located…What if, even though he could prevent it from happening, he were to permit some envious person to knock this pearl from his hand into the mire? And what if afterwards he were to take it from the mire and store it, still dirty and unwashed, in some clean and costly receptacle, intending to treasure it henceforth in this condition? Would you think him wise?

B. How could I think this? For would it not be much better for him to retain and safeguard, as clean, the pearl which is contaminated?

A. Would not God have been acting similarly? For God was holding in His own hand, as it were, man, who in Paradise was without sin and was to be placed in the company of angels; and God permitted the Devil, incited by envy, to thrust man (who nevertheless consented) into the mire of sin (for if God had willed to prevent the Devil, the Devil could not have tempted man). Would not, I say, God have been acting similarly if without there being any cleansing (i.e., without there being any satisfaction) He had brought man—stained by the mire of sin and going to remain for-ever in this condition—back at least to Paradise, from which he had been cast out?

The only act that would make satisfaction to God would be were man to restore his own nature to its original state of justice, where he could will justice for its own sake and love all created things for the sake of God.

One might think this leaves God waiting patiently for men to offer Him satisfaction, since the word signifies something that the offender gives to the offended. This image is wrong for two reasons: 1) God as the just governor of creation cannot be passive in restoring right order to His realm; 2) man is completely unable to give God satisfaction in any case. This second thesis is the more famous, so let us consider it first.
We have already seen in DLA 3-4 that man has the power only to preserve justice, not the power to acquire justice once it has been lost. So for man to return to justice, God would have to give him justice again, which without satisfaction would be the same as to be lenient, thus violating God’s rectitude.\(^8\) If we use the language of the honor, we can see another dimension to man’s dilemma. Man, by disobeying God, commits the sin of presumption, attempting to usurp God’s estate and place in society (he tries to ‘be like unto God’). To take away God’s estate is to reject God’s rule over creation and claim that rule oneself.\(^6\) But man is unable to rule over all creation—this is possible only for God. What’s more, after the fall he is so weak that he has to spend his whole life working to stay alive, in a now disordered creation, which makes it even less likely that he will stop choosing his own advantage in order to obey God’s will no matter what the cost. So clearly, restoring rectitude to all creation is beyond man’s ability, even were he able to restore rectitude to his own will, and even if God were not to have weakened man after his fall by making him corruptible and mortal.

Man has attempted to seize authority over creation, but he lacks the power to govern it responsibly. God has the power to govern, but he is constrained by justice from being lenient. As Creator, He also has the obligation to govern—as Justice itself, He

\(^8\) Elsewhere, Anselm predicts that even if man were to regain his original rectitude, he would for sure lose it again to the cares of this world. Cf. *De Conceptu Virginali et De Originali Peccato*, ch. 8. This text doesn’t conflict with DLA 3-4, since it is only a statement of probability—if man were restored to justice in a corrupted body, his body would soon enough drag him back into sin unless God’s grace intervened.

\(^6\) Cf. CDH I.12 (H&R): “this unfitness is so extensive that it makes injustice resemble God, for as God is subject to no one's law, neither would injustice be.” “Ad hoc etiam extenditur haec inconvenientia, ut iniustitiam deo similem faciat; quia sicut deus nullius legi subiacet, ita et iniustitia.”
ought to make creation as it ought to be. But He is again constrained by His decision to create men with freedom, which is part of human nature and thus cannot be taken away from man without introducing a contradiction. Anselm’s famous solution to this dilemma is that it is necessary for God to take on human nature, and thus restore it to its proper dignity. The God-Man retains His original justice, and thus on behalf of man gives God the honor He is due. The God-Man in addition accepts suffering and death, the punishment that is appropriate for sinful man but not required of the God-Man. This freely chosen gift, of infinite value because of the infinite value of the life offered, is satisfaction for any sin, at any time, on behalf of any human person the God-Man designates.

Here we have the solution to the problem of chapter 11 in the Proslogion. There, Anselm found himself at a loss to explain why God forgives some men and not others. Now Anselm has the tools to give an answer: Those who imitate the God-Man in giving themselves over to God “for the sake of justice” ("propter iustitiam"), who approach God

87 Obviously, Anselm does not mean necessary in the modern modal sense of “not possible that not.” Instead, he means a hypothetical necessity, in the sense that it is necessary for a perfect promise keeper to keep every promise he makes, though it is not necessary to make the promise in the first place.

88 Cf. CDH II.11

89 Cf. CDH II.10.

90 Cf. CDH II.14.

91 Cf. CDH II.16.

92 Cf. CDH II.19.
in the proper way,\textsuperscript{93} will be saved. Not everyone will accept this salvation, because some will refuse to submit to God, or are afraid of the suffering that will come. Anselm suggests that both prayer\textsuperscript{94} and suffering\textsuperscript{95} are necessary for the imitation of this God-Man.

That just people suffer has been the starting point of the modern problem of evil. Anselm’s response to this problem provides an interesting example of the way in which the God-Man anticipates the problems of living as one of the just in a disordered world, in an act of magnanimity:\textsuperscript{96}

A. Do you not realize that when He endured with patient kindness the injuries, the abuses, the crucifixion among thieves—which were all inflicted upon Him (as I said above) for the sake of the justice which He obediently kept—He gave men an example, in order that they would not, on account of any detriments they can experience, turn aside from the justice they owe to God? He would

---

\textsuperscript{93} Ibid.: “This is true—provided he approaches as he ought. Sacred Scripture everywhere teaches us how we are to approach the participation in such great grace and how we are to live under this grace.” “si accedit sicut oportet. Quemadmodum autem sit ad tantae gratiae participationem accedendum et quomodo sub illa vivendum, nos ubique sacra scriptura docet…”

\textsuperscript{94} Cf. CDH I. 24 and II.19. See also McIntyre’s treatment of this (\textit{St. Anselm and His Critics}, pp. 104-106).

\textsuperscript{95} Cf. DCD 21 (H&R): “because of the sin of our first parent human nature was made capable of suffering countless troubles; from this passibility grace works incorruptibility in us in many ways.” “Hominum enim natura propter peccatum primi parentis iam innumerabilium incommodorum facta est passibilis; ex qua passibilitate multis modis gratia nobis operatur incorruptibilitatem.”

\textsuperscript{96} CDH II.18 (H&R): A. An non intelligis quia, cum iniurias et contumelias et mortem crucis cum latronibus sibi, sicut supra diximus, propter iustitiam quam oboedienter servavat, illatas benigna patientia sustinuit, exemplum dedit hominibus, quatenus propter nulla incommode quae sentire possum, a iustitia quam deo debent declinent, quod minime dedisset, si secundum potentiam suam mortem pro tali causa illatam declinasset?
not at all have given this example if, as He was able to do, He had turned aside from the death that was inflicted upon Him for such a reason.

In the feudal arrangements of Anselm’s day, the most serious claim a vassal could make against his lord was the denial of justice. If true, it could release the vassal from his oaths of fealty to his lord.\(^9\) Anselm’s argument here is that those who imitate Christ will also have to give their lives for the sake of justice, and may have to experience injustice in this life for justice’s sake, but that the God-Man’s act of solidarity in suffering willingly ought to strengthen the affection and loyalty of His followers.

One might wonder why the fallen angels could not be saved by a similar means. Anselm emphatically states in CDH II.21 that they cannot be saved, but he presents an inadequate explanation why. There might be an explanation, however, in a cryptic passage in DCD 21 where Anselm suggests that man’s ability to suffer is a necessary precondition for his salvation.\(^9\) I think the idea is that angels, being incorporeal, cannot suffer except intellectually, and thus cannot be purified except intellectually, and they can only be punished intellectually by being separated from God. Were the Son to assume angelic nature as He assumed human nature, He would of necessity be a good angel, since He would have rectitude of will. Only because of man’s common descent from


\(^9\) DCD 21 (H&R): “because of the sin of our first parent human nature was made capable of suffering countless troubles; from this passibility grace works incorruptibility in us in many ways.” “Hominum enim natura propter peccatum primi parentis iam innumerabilium incommodorum facta est passibilis; ex qua passibilitate multis modis gratia nobis operatur incorruptibilitatem.”
Adam, and the consequent common fate of all men, is it possible for God’s assuming human nature to result in the needed solidarity with sinners; and only because human beings share the ability to suffer in the flesh without needing to be separated from God’s will is the God-Man able to suffer while remaining just. A body-soul makeup of human beings is a necessary condition for atonement, and so the fallen angels must be damned for eternity.

After Boso’s earlier despair in Book I, Anselm makes sure to point out that God’s free decision to restore rectitude to men is rightly called mercy.\footnote{Cf. CDH II.21.}

We have discovered that God's mercy—which, when we were examining God's justice and man's sin, seemed to you to perish—is so great and so harmonious with His justice that it cannot be conceived to be either greater or more just. Indeed, what can be thought to be more merciful than for God the Father to say to a sinner, condemned to eternal torments and having no way to redeem himself: “Receive my only begotten son and render him in place of yourself,” and for the Son to say “Take me and redeem yourself”? For the Father and the Son do make these respective statements, as it were, when they call and draw us to the Christian faith. And what is more just than that He to whom is given a reward greater than every debt should forgive every debt if it is presented to Him with due affection?

This is not the “suffering heart” of Isidore’s \textit{Etymologies}, nor is it the leniency proposed in CDH I.12, but it is a fair description of the action of God in saving man. The just

\footnote{CDH II.21 (H&R): Misericordiam vero dei quae tibi perire videbatur, cum iustitiam dei et peccatum hominis considerabamus, tam magnam tamque concordem iustitiae invenimus, ut nec maior nec iustior cogitari possit. Nempe quid misericordius intelligi valet, quam cum peccatori tormentis aeternis damnato et unde se redimat non habenti deus pater dicit: accipe unigenitum meum et da pro te; et ipse filius: tolle me et redime te? Quasi enim hoc dicunt, quando nos ad Christianam fidem vocant et trahunt. Quid etiam iustius, quam ut ille cui datur pretium maius omni debito, si debito datur affectu, dimittat omne debitum?}
action is the merciful action; restoring man to his proper dignity is both good for man and the way things ought to be. Anselm begins the dialogue assuming that God is merciful, and showing how justice, God’s responsibility for governing according rectitude, limits His ability to forgive. But God’s responsibility for governing with rectitude also necessitates that He restore as many souls as possible to justice. This responsibility leads Him not only to take on human nature, which otherwise would be beneath His dignity, but also to suffer for the sake of justice, and give human beings an example to follow so that they would not be discouraged at their own suffering.

It should be emphasized that for God mercy is not a pure perfection. This is because mercy can only be shown to beings who are not just. Since only creatures lack justice, were mercy an attribute identical with the divine essence, it would introduce into the divine essence a relation to creation. But that would make the act of creation necessary rather than free, which Anselm does not want to claim. Mercy then cannot be a pure perfection, but one might argue that the disposition to show mercy to fallen creatures would be a pure perfection (if we purify the idea of ‘disposition’ of any connotations inconsistent with immutability). Such a divine disposition to show mercy, which might be assimilated to divine love or divine goodness, would be revealed when mercy is shown. This is a classic use of mercy as a tool of statecraft—to show that the one showing mercy is kind and benevolent.

\[101 \text{ M 70; CDH 1.3}\]

\[102 \text{ M 74}\]
For Anselm, there is no difference between God’s merciful and His just actions. In the order of explanation, mercy seems to come first: where mercy is not mentioned, as in the *Monologion* and *De Casu Diaboli*, the vision of justice is straightforward (sinners are damned forever); but where mercy is mentioned (in the *Proslogion* and *Cur Deus Homo*), it enters into a complicated dialectic that enriches both concepts and arrives at equilibrium.

We have now given a fairly detailed picture of what Anselm has to say about justice in his speculative writings. If nothing else, we have established that justice is a persistent and important theme tying together almost all of his major works. Anselm is not often thought, even by theologians, to have paid sustained attention to the question of justice. We hope that this chapter has laid that misconception to rest.

We can also see how Anselm’s first conception of justice as obedient love of God had to be adjusted to make it compatible with mercy. We have argued that mercy is important for making justice humane, and so that any theory of justice ought to have something to say about mercy. Few theorists have been as thoroughgoing as Anselm in making these seemingly disparate concepts compatible, but if Jeffrie Murphy is right that Anselm’s paradoxes of mercy have not gone away, perhaps Anselm’s solutions to those paradoxes deserve more attention.

In the next chapter, we will look again at the *Cur Deus Homo*, this time in the process of examining Anselm’s political ideas. Anselm did not have a treatise on politics, but as we have seen, justice was a constant and deep concern of his. By surveying in more depth his various statements about politics (whether of God or man), we can
articulate a coherent and powerful political theory, one which puts the issues of punishment and mercy at its center. We argued that the failure to do this is a weakness in many political theories. Anselm’s decision to do this, we will argue, is a great strength of his theory. As we have seen in this chapter, Anselm’s theory of justice is quite abstract and formal. As we will see in chapter eight, it is this very abstract quality that makes it possible to develop an Anselmian liberalism for a pluralistic culture. Thus what for Anselm might be a weakness in his theory—its inability to give a rich answer to the question ‘what is the right thing to do now?’—will for our purposes be another strength.
CHAPTER SEVEN

ST. ANSELM’S THEORY OF MERCIFUL PUNISHMENT

7.1 Introduction

Since Rawls does not really have a theory of punishment, and Kant’s is so unacceptable, it seems that Robert Nozick and Jeremy Bentham might be the only two thinkers to have important theories of punishment that 1) are still considered viable and 2) are consistent with a larger moral and political theory that is still considered viable. I want to argue that St. Anselm’s view of punishment, if it is amended and developed in ways appropriate for democratic states, makes him eligible for this category. Neither Nozick’s libertarianism nor Bentham’s utilitarianism meets the conditions that Rawls lays out for a stable regime,¹ let alone a regime that protects the “most essential primary good,” self-respect.² On the other hand, the Anselmian view developed below is both stable and humane, or so I will argue.

In this chapter, we will look at Anselm’s theory of punishment and of the duty of a governor. Anselm’s view, unlike the others we have considered, assumes that questions about who to punish, and how much, are in the hands of particular people with the

¹ Cf. PL pp., xlii-xliii, 36-38, 65-6,140ff.
² PL p. 319.
responsibility of government, rather than an impersonal state criminal justice system. Thus, his theory of how and why to punish is also a theory about how to govern, which means that punishment plays a significant role in his view of the statesman’s responsibilities. Anselm actually develops two views of punishment, a preliminary view that requires strict justice, and his more developed view in which mercy plays the key role. As we will see, Anselm takes a certain view of divine providence as his model for earthly governance, again a move that is unusual among the theories of justice we have been investigating. One might expect this aspect of his view to make it harder to relate Anselm’s political theory to a modern secular nation state—the ways of God often being considered inscrutable and perhaps arbitrary—but surprisingly, as we’ll see in the last part of the chapter, it actually makes it easier. For Anselm holds that God is Justice itself, and Mercy itself, so that our theorizing about the actions of God is at the same time theorizing about justice and mercy. Jeffrie Murphy was thus right to think that the paradoxes of mercy taken from Anselm’s *Proslogion* were relevant to considerations of mercy in criminal sentencing. Indeed, recent historical scholarship into Anselm’s tenure as Archbishop of Canterbury shows that Anselm’s views about how to order the earthly realm are strikingly similar to the views of divine order we find in his philosophical and theological works. We will not have a chance to pursue these historical connections in any depth, but they do add credence to our revisionist political reading of Anselm’s views of justice and mercy.
7.2 Anselm’s Monastic Approach to Punishment

Anselm wrote and thought about punishment in several different places and contexts. As prior and later abbot of his monastery at Bec, he had the grave duty of correcting his monks according to the Rule of St. Benedict:³

In his instruction, the abbot should always observe the apostolic rule: “Reprove, entreat, rebuke” (1 Tim. 4:2). As the occasion requires, he should mix encouragement with reproof. He should show the sternness of a master and the love and affection of a father. He must reprove the unruly and undisciplined with severity, but he should exhort the obedient and patient for their own betterment. We warn him to reprove and punish the slothful and stubborn. He should not ignore the sins of offenders: but as soon as they appear and grow, he must root them out…He should verbally reprove the more virtuous and intelligent once or twice, but the stubborn, the proud, the disobedient, and the hard-hearted should be punished with whips, even at the first signs of sin. For “The fool is not corrected by words” (Prov. 29:19). And “Strike your son with rod and you shall deliver his soul from death” (Prov. 23:14).

According to the Rule, the abbot has to answer to God on Judgment Day not only for his own soul, but for the souls of the monks with whom he is charged.⁴ Hence there is a well-defined division of roles: each monk has to obey the abbot, and the abbot is responsible

---


⁴ Ibid.
for leading him to his salvation. The abbot is also responsible for soliciting the obedience of the monks, and correcting their disobedience. If some monk remains disobedient or persists in sin, the abbot is still responsible for him in God’s eyes unless he has done all that is within his power to win that monk over—a difficult standard to be sure one has met. In this, he lives up to his name, for ‘abbot’ is taken from Abba, or Father.

It was against the background of this grave duty that Anselm thought about punishment. How does one correct a disobedient monk so that he changes not just his behavior but also his desire to disobey? The monk has to obey in both mind and body, for if he merely conforms his outward behavior, he remains proud and in danger of losing his salvation. So the abbot has to punish with an eye towards winning the affection and obedience of the punished. Contrast this with the Machiavellian view of punishment. For both Anselm and Machiavelli, punishment is a key tool for governance. But for Anselm not only is it better to be loved than feared, it is a grave obligation to be obeyed out of love rather than out of fear.\(^5\) An abbot who ruled simply through fear would have failed to discharge the duties of his office, and on the Day of Judgment his soul might be in danger.

Anselm thus saw punishment as a tool that would take disorderly souls and order them to God. His concern in punishing is not unlike Rawls’s concern for stability—just as a regime is more stable if the citizens willingly uphold its principles as part of their good, so is monastic governance more stable if the monks willingly accept the commands of the abbot as indicating their path to holiness. If this analogy holds, it suggests that

\(^5\) Cf. ibid, ch. 5.
approaches to punishment which succeed in the monastery might have something to offer political theorists concerned about the stability of a regime. Indeed, we have reason to think that as Archbishop of Canterbury, Anselm himself tried to apply monastic principles of government to the governance of England.⁶

*Ex officio*, an abbot has to be concerned at all times with the good of each monk and with the common good of the monastery as a whole. He has some responsibilities for those outside the monastery (at least, those duties of charity and benevolence required of everyone), but for the most part these take second place to the well defined duties of his office. The diversity of personalities in any monastery ensures that balancing the good of each with the good of all requires great prudence and political skill, not least when it is necessary to punish some of the monks. For punishment not only must correct the individual monk and win back his obedience, it is also supposed to increase the order in the monastery and encourage the obedience and holiness of everyone. Punishment has to teach both the individual and the community, it has to defend the common good against the sins of the individual, it has to foster willing obedience in one and all, and it has to win the affection and love of all for all. (It is not surprising that the abbots most successful at this were often canonized.)

The right sort of punishment finds the mean between strictness and laxity. Strictness is the vice that overlooks the individual for the sake of the law; laxity overlooks the law for the sake of the individual. Both can be the result of a certain weakness on the part of a judge—the lax judge might lack the will to punish the offender, the strict judge might lack the capacity to forgive.

while the strict judge might lack the will to take an interest in him as a person. And both might end up undermining the goal they set out to defend: the indulgence of the lax judge might hurt the offender’s character, while the severity of the strict judge might undermine public confidence in the rule of law. Anselm’s view was that the prima facie contrast between strict ‘justice’ and lax ‘mercy’ misrepresented real justice and mercy, since the goal of justice and mercy properly understood is to preserve order in general and in the soul of each person. In the first part of his dialogue *Cur Deus Homo*, Anselm introduces his first theory of punishment in the course of an argument reductio ad absurdum, showing the trouble with understanding justice and mercy as contraries.

7.3  Anselm’s First Theory of Punishment and Governance: Strict Justice and the Demands of Order

In its rough outline, the reductio argument is quite simple. It takes place against the following background: the interlocutor Boso, arguing on behalf of the unbelievers, suggests that Christians detract from the dignity of God by suggesting He became man. Anselm responds that, on the contrary, the incarnation reflects God’s great mercy, thereby adding to His dignity. Boso grants that God must be surpassingly merciful, but argues that surely an omnipotent God could have found some more dignified way to save man. Anselm then begins his argument to show that the demands of justice necessitate that God assumed human nature to save man. Anselm argues ‘remoto Christo’, leaving

7 CDH I.3.

8 CDH I.10.
Christ aside.\(^9\) He then shows that without Christ, the demands of justice and order require that God punish all sinners strictly. Since all men are sinners, the demands of justice require that all men either restore the order that their sins disrupted, or they be punished. Since it is beyond the ability of man to restore the cosmos to its Edenic harmony, all men must be punished. If there were some being who was both God (and thus powerful enough to restore harmony to the cosmos) and yet also possessed human nature (and so could act on behalf of fallen men), then the demands of justice could be met without punishing all men. Since Anselm has stipulated that he is arguing *remoto Christo*, it follows that God cannot be merciful. But we have already agreed that God is merciful—a contradiction. Therefore, in order for God to be both just and merciful, He had to assume human nature. Here’s a reconstruction of the argument:

1) Assume *God is both just and merciful*
2) Assume *God did not assume human nature*
3) *Perfect justice* requires perfect order in the cosmos (*Def.*)
4) *Sin* is presumption, the deliberate usurpation of the perfect order (*Def.*)
5) All men are in a state of sin (*Def. Original Sin*)
6) Justice requires that, after sin, order be restored to the cosmos (3,4)
7) *Punishment* is the restoration of order by forcibly putting down the usurper (*Def.*)
8) For order to be restored, the sinner must either restore the order himself or be punished for his usurpation (4,6,7)
9) No creature can restore order to the whole cosmos (*Principle of transcendence*)
10) Since men are creatures, all men will be punished (5,8,9)

11) If God is merciful, then some men will not be punished (*Def.*)
12) Therefore, God is not merciful (10,11)

\(^9\) CDH Preface: (H&amp;R) “This book goes on to prove by rational necessity—Christ being removed from sight, as if there had never been anything known about Him—that no man can possibly be saved without Him.” “Ac tandem remoto Christo, quasi numquam aliquid fuerit de illo, probat rationibus ne nessariis esse impossibile illum hominem salvari sine illo.” See also CDH I.20.
13) Therefore, God is both merciful and not merciful (1,12)

For this argument to go through, Anselm has to defend his definition of sin, his concept of justice as order, and his claim that punishment will restore order. It is clear that order will be a key concept here.

The connection between strict punishment and stability, between law and order, is part of our conventional wisdom—a politician who wants to ‘get tough on crime’ proposes harsher punishments, not lighter ones. In chapter four, we saw how Rawls’s theory of punishment in an ideal society becomes very strict when the purpose of punishment shifts from promoting liberty to promoting stability. The same holds true for Anselm’s first view, which justifies a regime of strict justice in order to maintain perfect order. Anselm lays out this principle in ch. 12 of book I: “It is not fitting that God should ignore something that is disordered within His kingdom.” Anselm’s doctrine is almost always misread on this important point, so it will pay for us to examine it closely.

Anselm’s dominant image of God throughout Cur Deus Homo is as a good governor of an honor. In Anselm’s day, a man’s ‘honor’ was his estate. It consisted of significant holdings of land, and also the recognition by his vassals of his status as their

10 “Deum vero non decet aliquid inordinatum in suo regno dimittere.”

11 “The largest aristocratic estates, held by the earls and barons, were termed ‘honors’: assemblages of estates held by military tenure. They consisted not of unbroken stretches of magnate territory but of dozens of manors or holdings scattered across counties and intermingled with the lands of other lords…. The honor represented a principle of territorial administration different from that of the county. Rather than comprising a homogenous and continuous unit of land under the authority of one royal official, the honor was given whatever coherence it had by common tenure.” Robert Bartlett, England under the Norman and Angevin Kings, 1075-1225 (Oxford University
lord. In England immediately after the Norman Conquest (when Anselm was writing the *Cur Deus Homo*), it seems that this form of administration was at its peak; the Norman kings gave out choice pieces of the newly conquered lands to their allies, who in turn bestowed lands on their vassals, creating a connection of personal loyalty. Often the tenants in an honor would offer their lord personal military service or attend his court (called ‘the court of the honor’), where tenants would bring business before their lords, and on occasion, even bring suit against their lord. If the lord refused to deal with the legitimate grievances of his vassals, it was a serious charge. On the other hand, if a vassal tried to usurp the honor of a lord or failed to uphold the terms of his tenancy, he was committing a crime against the social bond that held together the kingdom, barony, dukedom, etc. The personal honor of a lord reflected upon his ability to protect his estate from such usurpation, and to require that the tenants uphold their obligations.

---

Press, 2000) p. 219. Bartlett points out that the kingdom of England was considered to be an honor.

12 “It was obviously always necessary for a vassal or tenant to know to whom he owed the services that formed a condition of his tenure….” Ibid. See also Southern, *Portrait*, p. 225: “The central feature of this estate was his landed property. But it also embraced his due place in the hierarchy of authority, his family background, and his personal honor.”

13 Bartlett writes (p. 219) “Much of England’s history depended on which bonds were to be more significant.”

14 Bartlett suggests (p. 221) that “denial of justice would have constituted at least an extenuating circumstance for even such a drastic step as imprisoning one’s lord.”

15 Cf. Bartlett, p. 221; Southern, *Portrait*, p. 226. Southern notes that after Anselm’s day the term ‘honor’ came to focus more on personal honor and less on the complex of political obligations. “The descent… was from a term associated with the structure of society to one denoting private feeling and reputation. Anselm’s references to
In his dialogue *De Veritate*, written before the *Cur Deus Homo* while Anselm was abbot at Bec, Anselm uses the word ‘rectitude’ or “rightness” to refer to the right ordering of all creation according to God’s providential plan. As we saw in the last chapter, he says that justice can be used as a synonym for rectitude, because something is just if it acts rightly according to its nature. Equating justice and rectitude leads Anselm to say some initially striking things. For example, he says a stone is just if it behaves as a stone ought, rather than, say, flying up in the air or singing in the rain. Anselm quickly explains that when we mean justice we usually apply it to human beings rather than rocks. Since human beings have a more complex nature, and justice is fulfilling their nature, justice for them is more complex—they not only have to behave rightly, they also have to act for the right reason, i.e., in accord with the right ordering of all creation by the Creator. Rectitude is an attribute of God, and so is identical with the divine nature. For God, to will justice is to will according to His own nature. Meanwhile, for man, to uphold justice is to will that the world be as God intends it to be, and to will this out of love of God.

Anselm was ordained Archbishop of Canterbury after writing *De Veritate* and before writing *Cur Deus Homo*. I want to suggest that Anselm’s exposure as archbishop to post-Conquest legal terminology allowed him to see the similarity between his own idea of rectitude and the honor of an Anglo-Norman lord. When Anselm used the term ‘honor,’ his audience heard a term that was more familiar than ‘rectitude,’ but which

God’s honor are to be interpreted in the earlier of these two contexts, which was that of his own day.” (*Portrait*, p. 226).
meant the same thing. Philosphically, rectitude as defined in *De Veritate* is the dominant idea here. But *rhetorically*, Anselm’s decision to use the word ‘honor’ has the effect of emphasizing that God is a governor of the universe, a king who tries to order His Kingdom well (which for God, means it is ordered perfectly).

This is important because Anselm’s language here has been criticized for being either too poetic (God is ‘insulted’ by sin) or too dependent on feudal legal terminology. But the idea of rectitude, which Anselm developed in his philosophical treatises (before he became Archbishop of Canterbury and grew so familiar with legal terminology), is the dominant idea here, as Anselm makes clear:

A. If angels and men always rendered to God what they ought to, then they would never sin….
A. Therefore, to sin is nothing other than not to render to God what is due….
A. The will of every rational creature ought to be subordinate to the will of God….

---

16 “God’s honor is the complex of service and worship which the whole Creation, animate and inanimate, in Heaven and earth, owes to the Creator, and which preserves everything in its due place. Regarded in this way, God’s honor is simply another word for the ordering of the universe in its due relationship to God.” (Southern, *Portrait*, p. 226)

17 See John McIntyre, *St. Anselm and His Critics* (Oliver and Boyd, 1954), in general, and especially ch. 2.

18 CDH I.11: (H&R, with some minor changes): ANSELMUS. Si angelus et homo semper redderet deo quod debet, numquam peccaret…. A. Non est itaque aliud peccare quam non reddere deo debitum…. A. Omnis voluntas rationalis creaturae subiecta debet esse voluntati dei…. A. Hoc est debitum quod debet angelus et homo deo, quod solvendo nullus peccat, et quod omnis qui non solvit peccat. Haec est iustitia sive rectitudo voluntatis, quae iustos facit sive rectos corde, id est voluntate. Hic est solus et totus honor, quem debemus deo et a nobis exigit deus. Sola namque talis voluntas opera facit placita deo, cum potest operari; et cum non potest, ipsa sola per se placet, quia nullum opus sine illa placet. Hunc honorem debitum qui deo non reddit, aufert deo quod suum est, et deum exhonorat; et hoc est peccare.
A. This is the duty which angels and men owe to God. No one who discharges this duty sins; and everyone who does not discharge it does sin. This is the justice or rectitude of will which makes men just, or upright, in heart (i.e., in will). This is the sole and complete honor which we owe to God and which God demands from us. For only such a will, when it is able to act, does works which are acceptable to God; and when it is not able to act, it alone is acceptable in itself, since without it no work is acceptable to God. Whoever does not give back to God this honor due Him removes from God what is His, and dishonors Him; and this is to sin.

This equation of sin with disobedience, theft, and disorder, and justice with rectitude and obedience, doesn’t require particularly feudal language.

On Anselm’s view, then, sin is not just disobedience, doing what one ought not to do. It is closer to an attempt at usurpation, of wresting from one’s lord the honor that is his right, and in general disrupting the governance of the honor. To resurrect the traditional term, it is presumption—asserting rights to a place in the honor to which one is not entitled. In Anselm’s day, if a vassal were attacked and the lord did not defend him, the relationship connecting the lord to that vassal and perhaps to all his vassals would be weakened. A lord who could not keep order to his honor would lose it, and the honor would descend into anarchy.

Cf. CDH I.15 (H&R): “When [a rational nature] wills what it ought, it honors God—not because it confers anything on Him but because it willingly submits itself to His will and governance. And, as best it can, it stays in its proper place in the universe and preserves the beauty of the universe. But when it does not will what it ought, then it dishonors God from its own point of view. For it does not willingly submit itself to His governance; and it disturbs (as much as lies in its power to do so) the order and the beauty of the universe—even though it does not at all injure or tarnish God's power or dignity [in and of itself].” “Quae cum vult quod debet, deum honorat; non quia illi aliquid confert, sed quia sponte se eius voluntati et dispositioni subdit, et in rerum universitate ordinem suum et eiusdemi universitatis pulchritudinem, quantum in ipsa est, servat. Cum vero non vult quod debet, deum, quantum ad illam pertinet, inhonorat, quoniam non se sponte subdit illius dispositioni, et universitatis ordinem et pulchritudinem, quantum in se est, perturbat, licet potestatem aut dignitatem dei nullatenus laedat aut decoloret.”
So when Anselm argues that God ought to uphold his honor, he is simply saying that he ought to preserve rectitude, the right order of creation. Throughout the *Cur Deus Homo*, he argues that it is not right that God should be dishonored—a statement, we can now see, that is practically a tautology. The use of the ‘honor’ language, though, does foreground God’s role as the good governor of creation, and this helps Anselm flesh out the idea of God’s justice by appealing to political intuitions about how one restores right relations after one party has offended another:\(^\text{20}\)

However, as long as [the sinner] does not restore what he has stolen, he remains guilty. But it is not enough for him merely to give back what has been stolen; rather, because of the wrong which has been inflicted, he ought to give back more than he has stolen. For example, if someone who injures another's health restores it, his doing so is insufficient payment unless he also gives some compensation for the painful wrong that was inflicted. Similarly, he who violates another's honor does not sufficiently restore this honor unless, in proportion to the injury caused by the dishonoring, he makes some restitution which is acceptable to the one whom he has dishonored. We must also note that when someone restores what he has unjustly stolen, he ought to return that which could not be exacted from him had he not stolen what belonged to another. Accordingly, then, everyone who sins ought to restore to God the honor which he has stolen. This constitutes the satisfaction which every sinner ought to make to God.

\(^{20}\) CDH I.11: *Quamdiu autem non solvit quod rapuit, manet in culpa. Nec sufficit solummodo reddere quod ablatum est, sed pro contumelia illata plus debet reddere quam abstulit. Sicut enim qui laedit salutem alterius, non sufficit si salutem restituit, nisi pro illata doloris iniuria recompenset aliquid: ita qui honorem alicuius violat non sufficit honorem reddere, si non secundum exhonorationis factam molestiam aliquid, quod placeat illi quem exhonoravit, restituit. Hoc quoque attendendum quia, cum aliquis quod iniuste abstulit solvit, hoc debet dare, quod ab illo non posset exigi, si alienum non rapuisset. Sic ergo debet omnis qui peccat, honorem deo quem rapuit solvere; et haec est satisfactio, quam omnis peccator deo debet facere.*

The term ‘satisfaction,’ which has its roots in Roman law, has brought more criticism on Anselm for his legalism. Again, see McIntyre, *Anselm and His Critics*. But as we can see, satisfaction is just the restoring of rectitude, returning things to the way they ought to be.
Hence, Anselm begins the main argument of the treatise by asking whether it is “fitting”
or “decorous” for God to dismiss man’s sins out of mercy alone, without asking for a
restitution of the honor taken from Him. Forgiveness here simply means not inflicting
punishment, i.e., being lenient. Anselm argues at length that this sort of leniency is a
violation of God’s justice, because it would lead to His kingdom being radically
disordered:

A…. Now, in the absence of satisfaction, to order sin rightly (recte) is just to punish it; therefore, if sin is not punished, something disordered is forgiven….
A. But it is not fitting that God should forgive something that is disordered within His kingdom….
A. Therefore, it is not fitting that God should forgive sin that goes thus unpunished.

Anselm’s concern is that God’s kingdom be ordered rightly, which requires that, as the
supremely just ruler, God should not be allowed to lose control of any part of creation, or
lose any of the obedience due to Him. Since nothing is greater than God, “then Supreme
Justice (which is identical with God Himself) keeps nothing more justly than God’s

____________________

21 CDH I.12 (H&R): “Redeamus et videamus utrum sola misericordia, sine omni solutione ablati sibi honoris debeat deum peccatum dimittere.

22 Ibid.: “Sic dimittere peccatum non est aliud quam non punire.” This evokes Anselm’s conjecture from DCD 23 that Satan, knowing God to be all merciful, assumed that God would not punish him for his disobedience. In that case too, the mistake was to think that mercy was identical with leniency.

23 Ibid.: “A…. Et quoniam recte ordinare peccatum sine satisfactione non est nisi punire: si non punitur, inordinatum dimittitur. A. Deum vero non decet aliquid inordinatum in suo regno dimittere. A. Non ergo decet deum peccatum sic impunitum dimittere.
honor in regard to the governance of things.” Given this requirement of justice, were God to allow His honor to be violated without punishment or satisfaction, it would show that He lacked the power to defend His honor (a silly thought) or that He wasn’t just (but we just said that He is Justice itself).

We can now see that punishment has a tight connection with justice conceived as right ordering. For punishment is much like a defensive war waged against a usurper who has threatened the right order. It is not something that a responsible ruler can forego if he wants to retain his position. To restore order, it is necessary not only for the lord to undo the material damage done by the offender, but also for him to reassert his authority over the presumptuous person by exacting an additional penalty. If he fails to do this, the challenge to his authority remains unmet, the security of all those who depend on the lord’s protection is uncertain, and the unchastened presumer might start drawing others to his camp. If the original order is legitimate and just, then the lord has an obligation in justice to preserve and protect it; any weakening of the order would be an injustice. Hence, in the absence of a satisfactory penalty voluntarily paid, punishment is necessary for a just order, and the failure to punish is unjust.

24 CDH I.13: “Si deo nihil maius aut melius, nihil iustius quam honorem illius servat in rerum dispositione summa iustitia, quae non est aliud quam ipse deus.”

25 The modern problem of evil can be seen as a form of this criticism—if God is a perfect governor, then how can there be such disorder in the world? Couched in these terms, though, the solution is easier to see: God allows the seeming disorder out of mercy and a respect for human freedom. See below.

26 Cf. CDH I.15 (H&R): “Indeed, the voluntary making of satisfaction for wickedness, and the demand for punishment of anyone who does not make satisfaction—to pass over the fact that God causes, in many ways, good things to come from evil things—retain, in the universe, their proper place and preserve the beauty of its
not tolerate any deviation from perfection, however slight, and so each act of disobedience must be punished, unless satisfaction can be made.

The reader might be wondering, though, how anyone could be thought to usurp God’s reign. Is there any doubt that an omnipotent Ruler would retain His throne no matter what, that no creature could usurp her own Creator? Anselm agreed that such usurpation is impossible. Indeed it’s because He is omnipotent that God is able to be merciful despite the demands of maintaining perfect order. But Anselm also held that a free and rational creature could be disruptive of the divine order, not by dethroning God, but by denying Him the relations of obedience and love that He deserves.²⁷

order. If where wickedness tries to disturb right-order Divine Wisdom did not include these things [i.e., did not provide for the making of satisfaction and the exacting of punishment], then in the universe (which God ought to order) there would occur a certain marring as a result of the violation of the order’s beauty; and God would seem to fail in His governance. Just as these two results are unfitting, so they are impossible; therefore either satisfaction or punishment must follow upon every sin.” “Ipsa namque perversitatis spontanea satisfactio vel a non satisfaciente poenae exactio--excepto hoc quia deus de malis multis modis bona facit - in eadem universitate suum tenent locum et ordinis pulchritudinem. Quas si divina sapientia, ubi perversitas rectum ordinem perturbare nittur, non adderet, fieret in ipsa universitate quam deus debet ordinare, quaedam ex violata ordinis pulchritudine deformitas, et deus in sua dispositione videretur deficerre. Quae duo quoniam sicut sunt inconvenientia, ita sunt impossibilia, necesses est ut omne peccatum satisfactio aut poena sequatur.”

²⁷ CDH I.15 (H&R): A. Dei honori nequit aliquid, quantum ad illum pertinet, addi vel minui. Idem namque ipse sibi est honor incorruptibilis et nullo modo mutabilis. Verum quando unaquaque creatura suum et quasi sibi praeceptum ordinem sive naturaliter sive rationabiliter servat, deo oboedire et eum honorare dicitur, et hoc maxime rationalis natura, cui datum est intelligere quid debeat. Quae cum vult quod debet, deum honorat; non quia illi aliquid conferit, sed quia sponte se eius voluntati et dispositioni subdit, et in rerum universitate ordinem suum et eiusdem universitatis pulchritudinem, quantum in ipsa est, servat. Cum vero non vult quod debet, deum, quantum ad illam pertinet, inhonoret, quoniam non se sponte subdit illius dispositioni, et universitatis ordinem et pulchritudinem, quantum in se est, perturbat, licet potestatem aut dignitatem dei nullatenus laedat aut decoloret.
Nothing can be added to or subtracted from His honor, considered in itself. For His honor is, in itself, incorruptible and altogether immutable. But when each single creature keeps, either by nature or by reason, its proper place [in the order of things]—a place prescribed for it, so to speak—it is said to obey God and to honor Him. And this [holds true especially for] a rational nature, which has the gift of understanding what it ought to do. When a rational nature wills what it ought, it honors God—not because it confers anything on Him but because it willingly submits itself to His will and governance. And, as best it can, it stays in its proper place in the universe and preserves the beauty of the universe. But when it does not will what it ought, then it dishonors God from its own point of view. For it does not willingly submit itself to His governance; and it disturbs (as much as lies in its power to do so) the order and the beauty of the universe—even though it does not at all injure or tarnish God's power or dignity [in and of itself].

A disobedient will is in itself an impermissible disorder in God’s perfect kingdom, a marring of its perfect beauty. So the tiniest disobedience is a disorder. Anselm defines obedience as “when a rational nature, not by necessity, but freely, preserves a desire that it has received from God,” and he believes that the obedience of free creatures adds to the beauty of the universe “most of all” (maxime).

So Anselm’s idea that justice is right ordering, and sin is the disrupting of that right ordering through presumption, leads him to conclude that God cannot be lenient to sinners if He is to be perfectly just. This view is roughly as strict as Kant’s view, which we examined earlier, but for slightly different reasons. Anselm’s strict justice is driven by political necessity rather than regard for the law. There is no pretense that all actors are equal and autonomous in Anselm’s view—one person is in charge of the others, and the responsibility for ensuring right order falls to that person. It does not matter if a failed

______________________________

28 CDH I.10: “Nam tunc est simplex et vera oboedientia, cum rationalis natura non necessitate, sed sponte servat voluntatem a deo acceptam.”
governor is conscientious and benevolent or culpably negligent—if he fails to discharge his responsibilities, he is guilty of an injustice. Even if he fails out of weakness, he is guilty of having failed to acquire the power necessary to maintain a just order. Just as authority is distributed unequally, so is responsibility distributed unequally. The governors are judged as harshly as anyone, and perhaps more so, since considerations of mens rea are not always relevant to their judgment.

7.4 Anselm’s Second Theory of Punishment: Restoring Order by Merciful Solicitude

When he moves to his second theory of punishment, Anselm does not abandon his high standards for rulers, but he does allow mercy for the ruled. How he does this is rather complex and controversial, in part because it is so often approached from the wrong angle. I hope that after showing in the last section the conception of political order behind Anselm’s account of strict justice, in this section we can recover the political role he posits for mercy.29

Described in crude strokes, as it often is, Anselm’s argument for God’s mercy hardly seems to be such at all. Indeed, it is often criticized for depicting a God without mercy. For the way that God shows mercy, in Anselm’s scheme, is by becoming incarnate and then sacrificing himself in order to pay the debt for human sin. Since the God-Man possesses human nature, He is able to accept the penalty for human sin on

29 The sense in which our topic is political order might not be obvious, since we have been discussing God’s order over the universe. But since we argued in chapter five that the pure perfections are defined so as to apply to creatures as well as the Creator, and we showed in chapter six that justice and rectitude were pure perfections, we think we are entitled to read Anselm’s analysis of God as governor as applicable to human governors, mutatis mutandis.
man’s behalf. Since He is God, he is able to offer to God not only His complete 
obedience, which is what all men owe God in the first place, but an infinite payment in 
addition. By dying when He didn’t have to (because He didn’t sin), the God-Man gives to 
God the free gift of His life, which, as divine, is infinitely valuable. With His complete 
obedience and infinite payment, the God-Man is able to restore honor to God no matter 
how many sins man commits. Thus, the penalty for sin is satisfied, God’s thirst for justice 
is quenched by the blood of His own Son, and human beings can be admitted to heaven.

The critics of this argument tend to reject what they suppose to be Anselm’s 
image of God as jilted, wrathful, and eager for vengeance, a representative of a 
Mediterranean ‘honor’ culture who has to exact retribution, a Shylock who won’t show 
mercy unless repaid all that He is owed. The political reading of honor, order, and sin 
that we proposed in the last section shows that Anselm is able to avoid most of these 
criticisms. Sin is an act of presumption against God’s rule over the universe, and it needs 
to be put down so that the universe can be perfectly ordered again. Because the standards 
of order are so high—a perfect King should have a perfectly ordered Kingdom—God 
cannot simply dismiss the punishment, any more than a king could stand by while his 
authority is undermined and his kingdom is usurped. As we suggested above, punishment 
on this view is similar to a defensive war, though on a smaller scale. By understanding 

30 The criticisms have been basically the same for centuries. Again, see McIntyre, 
St. Anselm and His Critics.

31 The connection between war and punishment seems to be deep, perhaps 
because they are both species of justified violence or harm to another on behalf of 
society. H. L. A. Hart even develops something like a jus ad bellum/jus in bello 
distinction when he argues that it is one thing to justify a system of punishment, another
Anselm’s argument as about political order rather than as a commercial tort, it becomes evident that punishment can be seen as a necessary means of self-defense, while on the commercial view it is never quite clear what would be the harm if God were to forgive the debt.\textsuperscript{32}

to justify the distribution of punishment within that system. See \textit{Punishment and Responsibility}, ch. 1.

\textsuperscript{32} See William Kneale, “The Responsibility of Criminals” in Acton, \textit{The Philosophy of Punishment}, pp. 175-183, for a helpful discussion of the influence of economic metaphors on the idea of retribution, and of some of the drawbacks of this metaphor. Kneale claims that unless we are prepared to rehabilitate the very old connection between the moral word ‘ought’ and the commercial word ‘owe,’ we should be suspicious of the whole economic metaphor of moral obligation. “In many cases of hurt it is already very difficult to decide what is appropriate reparation, because there can be no restoration of the situation as it was before the wrongdoing; but at least we know what considerations are relevant, namely the nature and amount of suffering to be alleviated. When, however, we pass from compensation in a strict sense to the notion of expiating a sin or nullifying it by suffering, we have no standard by which to determine what suffering is appropriate to the sinner. For the old maxim of an eye for an eye is obviously inapplicable except in a very small range of cases. How, for example, can a childless youth who curses his mother have precisely the same done to him? In short, the whole theory of retribution seems to arise from a muddle-headed attempt to extend the notion of reparation beyond the limits within which it makes good sense” (p. 182).

Indeed, the commercial view has substantially less textual support than is usually thought. In CDH II.18, Anselm explicitly denies that, when he speaks of a \textit{debitum}, it should be taken to imply that there is a \textit{debito} or financial debt. In CDH I.11, in which Anselm lays out his view of sin and of satisfaction for sin, Anselm uses the constructions “\textit{debitum solvere},” “\textit{solvit quod rapuit},” and “\textit{solvit quod abstulit}.” The image of the last two is of releasing what one has seized and carried away, while the first is a phrase that can mean ‘discharge a duty’ just as easily as it can mean ‘pay back a debt.’ (In II.17, Anselm uses a parallel phrase, “\textit{promissum solvere},” to mean ‘to fulfill a promise’ which is closer to discharging a duty than paying a debt.) Since the vision of sin for Anselm is as an unjust taking of what is rightfully God’s (see also DIV 10), it makes much more sense to avoid the commercial translations of ‘\textit{solvere}’ altogether. Perhaps ‘resolve a debt’ might be an appropriately ambiguous translation of “\textit{debitum solvere},” since one can resolve both commercial and noncommercial debts. The same goes for ‘\textit{reddere}’ and ‘\textit{dare}’ which are Anselm’s preferred terms for describing the something more that one should give to God as a penalty.
Anselm holds that a kingdom would be well ordered were all its guilty citizens to be punished. In that case, all would recognize the king’s authority, and every material harm would be restored as much as possible. But he does recognize that this view has an odd consequence—namely if all the citizens were to be guilty of crimes that demanded infinite punishment, the kingdom would be rightly ordered only if all the citizens were punished forever. That does not seem a satisfactory definition of ‘rightly ordered.’ It does not seem to promote what Rawls rightly sees as an important aspect of stability, namely an affection for the principles of the regime among future generations. Surely, as Anselm asks, “it is necessary that some human beings should attain happiness,” isn’t it? Especially when the King is responsible for creating his realm in the first place. If God created human beings so that they could enjoy eternal happiness, and if none were to reach such happiness but rather all were to be punished eternally, then He would have created either in vain or out of malice. Therefore, once God created mankind for the purpose of enjoying beatitude, He committed Himself to ensuring that at least some people should enjoy beatitude. Which means that He needs a way of reconciling the need for keeping order (including the order within His subject’s souls) with the need for some guilty people to enjoy eternal happiness.

There is one further obstacle to God’s perfect governance—human freedom. Anselm thought that since God was the Author of human freedom, He had a particularly strong stake in defending that freedom. On Anselm’s view, God, in creating free, rational beings, made something just beyond His control: He can destroy them, He can make them fear Him, but He cannot force them to love Him. If those rational beings are to be

33 CDH I.25: “A. An non intelligis ex iis quae supra diximus, quia necesse est aliquos homines ad beatitudinem pervenire?”

34 Ibid.
true to their nature, they must worship God freely—only that sort of worship is in accord with justice. Anselm believes that God could not simply force man to be just, whether by zapping his will and infusing justice, or by overpowering his will by manifesting the glories of heaven or the horrors of hell. Simply to make man’s will just would require God to violate his own decision to create man with free choice of the will, and such vacillation would not be appropriate in God. On the other hand, if God tries to bribe man into obedience by manifesting His Glory, or cow him into submission through visions of hell, man would obey God for selfish reasons rather than out of love. So God is forced to teach man and try to persuade him, rather than exercise His power to coerce obedience. The force of this requirement becomes clear if we introduce more of Anselm’s philosophical machinery.

Anselm defines justice as *rectitude*, which is the right ordering of all things in the universe, so that all things inhabit their God-given place. If a stone fulfills its nature, say by not floating up into the sky, it is a just stone. It *acts* according to rectitude. Since animals don’t just act but also have *wills*, a horse that willingly goes to the horse pasture is a just horse. Human nature is more complex still, so justice for human beings is also more complex. Since man has faculties for acting, willing, and *knowing*, justice consists in acting and willing rightly for the right reason. Anselm thinks that he can express this

---

35 Cf. DCD 13-14.

36 DV 12.

37 Justice for a stone and for a horse are less than proper senses of justice for Anselm. They follow from his identification of rectitude with justice, but they don’t accord with normal ways of speaking.
definition of justice more briefly: a rational being is just if its will is just. That’s because a rational being can will rightly only if it knows what is right, and it can act rightly only if it wills rightly. The will is just if it fulfills the purpose for which God made it, namely, to will rightly, to will what God intended. There is an added complexity in human willing, since, as Anselm notes, “One can will to do what one ought, but out of a desire for vainglory or fear of punishment,” or even because forced to by another, rather than because it is the right thing to do. So justice for human beings must be not only willing the right thing, but willing the right thing for the right reason, for no other reason than for the sake of rectitude itself. For rational beings, then justice is “rectitude of the will preserved for its own sake.” Or as he put it in the Monologion, “the purpose [of the rational creature] is to love the supreme essence and only to love other things for the sake of the supreme essence.” Thus, rectitude doesn’t just mean the absence of disorder, the subjugation of all usurpers. It has positive content—that God be loved for His own sake.

In his dialogue On the Fall of the Devil, Anselm argues that there are two powers of the uncorrupted will, one that preserves justice and the other that seeks what is advantageous. The decision to sin is the decision to seek one’s advantage without

38 DV 12. “Qui autem non nisi coactus aut extranea mercede conductus vult quod debet: si servare dicendus est rectitudinem, non eam servat propter ipsam sed propter aliud.”

39 Ibid. “Iustitia igitur est rectitudo voluntatis propter se servata.”

40 M 68: “Nihil igitur apertiis quam rationalem creaturam ad hoc esse factam, ut summam essentiam amet super omnia bona,…immo ut nihil amet nisi illam aut propter illam.”

41 DCD 12-13.
desiring to preserve justice—to seek what seems good without regard to God’s right ordering of the cosmos. Sin always reflects a miscalculation, since it is always more advantageous to will what God wants us to will. We always pursue what seems to be our own advantage, but sometimes we choose what merely seems to be to our advantage rather than what actually is to our advantage.

Anselm defines freedom as the power to preserve justice understood in this sense. If I can only pursue my advantage, then I am at the mercy of my environs. If you convince me that doing action A is to my advantage, then on Anselm’s view I have to choose action A for as long as I remain convinced. I am a slave to what seems advantageous. If, on the other hand, I also have the ability to preserve justice, then I have a choice: I can do what is just, or I can do what seems to be my advantage. I can be told that doing what is just is always to my advantage, but unless I am convinced of that fact, I am still able to do otherwise. So only if I have the ability to act justly, do I really have a choice about whether to choose what is advantageous. Anselm’s view allows that there will be all sorts of deliberation about say, whether Rocky Road or Cookies & Cream is more advantageous for me now. Yet, if ever we become certain that something is absolutely to our advantage, we cannot help but choose it. What’s more, since we cannot help but choose it, we are not praiseworthy for choosing it.

Anselm argues that (creaturely) justice cannot be acquired but can only be given. Since God gives rectitude of will so that the creature may keep it, when the creature

---

42 DCD 16, 27.
doesn’t keep it, it’s lost for good. Anselm argues that freedom is a power or ability, and freedom of will is the ability to do what the will ought to do, which is to preserve rectitude for its own sake. This ability is part of human nature. Even sinners, who lack both rectitude of will and the ability to acquire justice on their own, nonetheless always have the power to keep rectitude if it is restored to them. Nothing can take away the power to preserve justice—someone can be bound unwillingly and killed unwillingly, but one cannot will unwillingly, and so we are always responsible for what we will and how. Thus, free will cannot be overcome by another power, by strong temptations, or even by God.

So when it comes to soliciting the obedience and love of fallen human beings, Anselm’s God is in a complicated position. God cannot restore justice to a person’s will if it is indubitably clear that it is to the person’s advantage to obey Him, since it is possible to obey God out of advantage rather than justice. One might follow His commands out of fear of punishment, perhaps, or desire for the advantages of heaven. On Anselm’s view, such obedience would be neither free nor praiseworthy. More

43 DLA 3. Note that there are not degrees of justice on his view.

44 DLA 3: “omnis libertas est potestas”

45 DLA 4.

46 DLA 5.

47 DLA 6-7.

48 DLA 8. Anselm thinks that a will can be said to be a “slave to sin” (John 8:34) because it cannot acquire rectitude on its own, but since it never loses the power by which it could preserve rectitude were it to reacquire it, it is still free.
importantly, the person would not have a just will, and therefore would not truly be
converted to God’s party. If God were to make manifest the horrors of Hell or the glory
of Heaven, He could force someone to conform to His commands. But if He wants to
restore her to justice, He has to persuade her that justice is worth desiring for its own
sake. And since justice for rational beings is to love God above all and to love all things
only for God’s sake, He has to persuade her to love Him and obey from that love.

We’ve now seen all the constraints Anselm thinks God has placed on His saving activity:

A. All people must be just to be saved. Justice is to act as God wills out of love
   for God above all else. It is also characterized as perfect voluntary obedience
to God’s will.
B. A person who on at least one occasion has pursued her own apparent
   advantage rather than acted from love of God loses the ability to act from love
   of God and cannot regain it under her own power.
C. As a consequence of original sin, no human being acts from love of God
   above all else.
D. Perfect rectitude or order requires that all those who disobey God’s will (usurp
   His authority/ prefer their own seeming advantage) be punished unless (i) they
   repair whatever damage they’ve done, (ii) in the future follow God’s will out
   of perfect voluntary obedience, and (iii) in addition pay some appropriate
   penalty.
E. No mere creature can make satisfaction for disobeying God.
F. Some people must be saved.
G. It is impossible to force a free and rational person to act out of love for God
   above all else.
H. It is possible to persuade a free and rational person to try to love God above
   else, even by appealing to her own advantage (at least in part), so long as the
   appeal is not made overpowering.

In the above summary, A-E are a restatement of the first theory of punishment, while F-H
are additional considerations. Anselm does not repudiate what he said earlier, but he does
add the important stipulation that some people must be saved. Anselm declares that God
shows His mercy by saving man from punishment and restoring him to the beatitude that he had lost.\footnote{CDH I.3 (H&R): “A. We do not at all dishonor or affront God; instead, giving Him thanks from our whole heart, we laud and proclaim the ineffable depth of His mercy. For the more miraculously and wondrously He has restored us from such grave and such deserved evils in which we found ourselves—restored us to such great and such undeserved goods which we had lost—the more He has shown love and graciousness toward us.” “A. Nos non facimus deo iniuriam ullam aut contumeliam, sed toto corde gratias agentes laudamus et praedicamus ineffabilem altitudinem misericordiae illius, quia quanto nos mirabiliaus et praeter opinionem de tantis et tam debitis malis in quibus eramus, ad tanta et tam indebita bona quae perdideramus, restituit, tanto maiorem dilectionem erga nos et pietatem monstravit.”}

Although Anselm does not explicitly define mercy as he does justice, we can suggest the following definition as plausible and consistent with his view of justice: mercy is the restoring of rectitude to the will of another for the sake of rectitude itself. One is not merciful to oneself, but only to another. Nor is one merciful to another for whom all is as it should be, but only when there is something amiss. Anselm’s view is that mercy does not lead to disorder, but is a restoration (\textit{restauratio})\footnote{Ibid.} of the ability of a rational being to pursue her own natural end (i.e., loving God above all else), which is justice and rectitude for her. This restoration of rectitude in another must itself be in accordance with justice, and so it must be for the sake of rectitude and not for some injustice.\footnote{As in \textit{Measure for Measure} II.4, where Angelo offers ‘mercy’ to Claudio on the condition that Isabella sin with him.} It should be clear that one cannot show mercy unless one has the power to restore the other to justice, though we might say that someone is \textit{merciful} who wants to show mercy and who would restore the other to rectitude if possible. (This distinction

\textit{...}
between the disposition to show mercy and the power to effect mercy will be important for our discussion later, when we will tie it to what we said in chapter two about the connection between power and responsibility.) It also follows that being merciful is a part of justice, since it is acting in accord with rectitude for the sake of rectitude itself.\(^{52}\)

We will need to legislate a bit of vocabulary here: Given human freedom, we cannot force someone else to be just, i.e., to have rectitude of will for the sake of rectitude itself. Likewise, we cannot restore her to justice against her will. So we can offer mercy or show mercy to someone, but it is always possible for her to reject our overtures and thus reject our mercy. When we offer mercy we offer to another the restoration of rectitude to her will for the sake of rectitude itself. If we want to unpack this definition a little more, we might say that mercy is the mitigating by the appropriate authority of a threatened punishment for the reason that the wrongdoer now has (or in the considered view of the authority is likely to have) a just will and is no longer likely to be a source of danger or disorder.

If this definition is adequate, then we can say that, on Anselm’s account, God must show mercy to at least some human beings, as a consequence of His act of creation and the subsequent Fall. God’s own Supreme Justice demands this. Since human beings cannot grant themselves mercy, God who is both merciful and omnipotent must do it Himself.\(^{53}\)

\(^{52}\) Cf. M 78, in which Anselm argues that to love Supreme Justice is to have a practical desire to effect justice in the world and to alleviate all injustices, given the opportunity.

\(^{53}\) CDH I.4 (H&R): “A. Do not the following considerations seem to constitute a very cogent argument for why God ought to have done those things about which we are
Anselm considers another possibility, that God could save man through either an angel or by a man that He created without sin. Anselm rejects this proposal, in a way that is useful for our political reading of the *Cur Deus Homo*. He has a two-part objection to this. First, if man were not saved by a divine person, he “would rightly be deemed to be the servant of whatever other person would redeem him from eternal death.”

We might find this outcome objectionable in itself (what liberal democrat wants to be a servant?), but for Anselm and his audience being in a relation of obedience, servitude, and debt to a patron or lord was quite normal. The second part of the objection, though, is more significant. Before the Fall, human beings were the servants of God alone. But if they were to be saved by someone other than God, then that person would be their immediate lord, and they would no longer serve God directly. In that circumstance, they would not be restored to their original place in God’s order, and they would owe love and obedience to their immediate lord, rather than exclusively to God. Indeed, since the non-divine person would be a mediator between God and those whom he had saved, the non-divine savior would assume responsibility for those he had saved (much as we saw an abbot did). God would deal only with the mediator, who would then rule his servants. The people ‘saved’ in this way could not be happy, since they lack the love of God for His speaking?: viz., that the human race—His very precious work—had utterly perished; and it was not fitting that God's plan for man should be completely thwarted; and this plan of God's could not be carried out unless the human race was set free by its very Creator.”

“A. Nonne satis necessaria ratio videtur, cur deus ea quae dicimus facere debuerit: quia genus humanum, tam scilicet pretiosum opus eius, omnino perierat, nec decebat ut, quod deus de homine proposuerat, penitus annihilaretur, nec idem eius propositum ad effectum duci poterat, nisi genus hominum ab ipso creatore suo liberaretur”

54 CDH I.5 (H&R): “A. An non intelligis quia, quaecumque alia persona hominem a morte aeterna redimeret, eius servus idem homo recte iudicaret?”
own sake, something they should have had.\textsuperscript{55} And therefore, they would be eternally unhappy.

On the other hand, if God Himself is the ‘mediator,’ then there is a perfect alignment of the servitude one owes to the mediator as patron and the obedience one owes in justice to God. To love the mediator is to love God; nothing is lacking in the former love. Man is restored to his proper place in the divine order, and restored to his previous dignity—he has no king but God. Since, as we saw above, this mediator has to be human in order to represent human beings,\textsuperscript{56} and divine in order to be able to effect mercy, Anselm concludes, \textit{sola ratione},\textsuperscript{57} that salvation occurs, necessarily, through a \textit{Deus Homo}, a God-Man.

The God-Man is not only merciful, and so reluctant to punish,\textsuperscript{58} but is also sufficiently powerful that He can \textit{effect} mercy, by making satisfaction for man’s sins. Anselm notes that the God-Man must be powerful because only He is able to give gifts to God. Because the God-Man is God, He is able to preserve justice by willing whatever

\begin{itemize}
\item \textsuperscript{55} CDH I.24 (H&R): “happiness is a state of sufficiency in which nothing needed is lacking,” “beatitudo est sufficientia in qua nulla est indigentia.”
\item \textsuperscript{56} In I.22, as one of many arguments in behalf of his conclusion, Anselm argues that man was originally a mediator between God and the devil, acting \textit{pro Deo} to show the devil that even such a weak creature as man could preserve justice. In this man would have conquered the devil and given honor to God. Anselm concludes that the savior would have to assume this role as mediator acting on God’s behalf to conquer the devil. So the mediator would have to represent both human beings and God.
\item \textsuperscript{57} CDH I.20.
\item \textsuperscript{58} In CDH I.7, Anselm distinguishes between the devil, who out of malice wishes to punish others, and someone who punishes reluctantly out of a love of justice. This quick distinction is important. Anselm clearly does not view God as vengeful and angry, but instead views Him as more like a father—or an abbot.
\end{itemize}
God does. And furthermore, because He is human, He is able to offer to God His suffering and His own life. Anselm believes that death is a consequence of sin, so that God would not demand death from a sinless human being.\(^5^9\) Therefore, when the God-Man offers the suffering and death of His human body to God, it is not the fulfillment of an obligation, but a gift that adds infinitely to God’s honor. In return for the gift of honor, God is able to grant the God-Man’s request (which is also His own will) that all those who ally themselves with the God-Man be saved.\(^6^0\) Because the addition to God’s honor by the gift of the God-Man is so great, mercy for those loyal to the God-Man does not lead to disorder in the universe.

This solution to God’s political problem, though, still leaves unfinished one very important practical task. The God-Man has the metaphysical power to restore the honor taken by any sinful human being past, present, and future. But He runs up against the limits that He set to His own power in creating human beings with freedom. He needs sinful people to want to serve God as He wants them to serve Him, and to want this for the sake of rectitude itself, rather than for their own seeming advantage. As we saw earlier, He cannot convince people to be just by overwhelming them with the disadvantages of hell or the advantages of heaven, since those who choose to obey God because it is to their own obvious advantage are not obeying out of love for rectitude itself. So He has to bring people to the point that they want to follow Him even when at times it seems to their disadvantage. To put this in terms of a distinction we developed in

\(^5^9\) CDH I.9; II.11.

\(^6^0\) CDH II.19.
chapter one, He needs them to hold the principles of justice as personally normative, so that they follow them no matter what their reason and senses tell them is to their personal advantage. What’s more, the ‘principles of justice’ to which they adhere cannot be transparent or rationally discernable to all people, because that would make them seem advantageous. They need to trust that the God-Man is their lord and their mediator, and they need to offer to be His servants, even though He is unable to prove to them that He is the God-Man without destroying the possibility of their becoming just. So He needs to persuade them to give up their own wills and trust themselves to the benign heteronomy of another, just as the abbot has to persuade recalcitrant monks to submit to his rule.

Solidarity and loyalty, rather than autonomy and reason, are necessary in circumstances of benign heteronomy. Thus, the God-Man is faced with the delicate task of trying to win the personal loyalty of a people who since the Fall had been living only for their own advantage.

One way to win solidarity is to show solidarity. In a brief departure from his argument *sola ratione*, Anselm explains the passage in Philippians 2:6-9, in which Christ is seen to abandon His divinity to take up humanity in order to be obedient unto death. Christ’s obedience, Anselm argues, was so great that He put aside the seeming advantage of His human nature out of love of justice:  

However, where He says “Father, if it be possible, let this chalice pass from me; nevertheless not as I will but as You will,” (Matt. 26:39) and “If this chalice cannot pass from me except I drink of it,  

---

61 CDH I.9 (H&R): Ubi autem dicit: »pater, si possibile est, transeat a me calix iste; verumtamen non sicut ego volo, sed sicut tu«; et: »si non potest hic calix transire, nisi bibam illum, fiat voluntas tua«: naturalem salutis per voluntatem suam significat appetitum, quo humana caro dolorem mortis fugiebat.
Your will be done”: (Matt. 26:42) by the reference to His own will He signifies His natural desire for security—a desire in terms of which His human flesh shunned the pain of dying.

And again:62

Since each man's will draws or moves him toward that which he wills unwaveringly, God is not inappropriately said to draw him or to move him when He gives him such a will. By this drawing or moving, no compelling force (violentiae necessitas) is signified; instead, there is signified a voluntary and devoted retaining of the good will which has been received. Therefore, if in this respect we cannot deny that the Father drew and moved the Son toward death when He gave Him that willingness [to die], who cannot see that in this respect the Father gave Him the ‘command’ so that he underwent death voluntarily, and gave Him the chalice of which He drank willingly?

These passages establish the solidarity of Christ with man in two ways. God is supposed to have respected Christ’s human freedom by asking for His obedience and not violently extracting it. Thus, Christ’s obedience was free. And secondly, the obedience was against the grain of His human nature, which feared death and naturally sought its own advantage. Anselm notes both characteristics to emphasize the similarities between the God-Man and other human beings.

By emphasizing the humility and solidarity of the God-Man, Anselm creates the image of a different kind of feudal lord, one who shrinks the distance between master and

62 CDH I.10 (H&R, with minor changes): Quoniam namque voluntate quisque ad id quod indeclinabiliter vult, trahitur vel impellitur, non inconvenienter trahere aut impellere deus, cum talem dat voluntatem, affirmatur. In quo tractu vel impulsu nulla intelligitur violentiae necessitas, sed acceptae bonae voluntatis spontanea et amata tenacitas. Si ergo hoc modo negari nequit patrem voluntatem illam dando ad mortem filium traxisse vel impulsisse: quis non videat eadem ratione »mandatum« illi ut mortem sponte sustineret, et »calicem« quem non invitus biberet, dedisse?
servant and shares his servants’ struggles and fears, indeed, one who seeks to be imitated as an example.\textsuperscript{63}

How could He give Himself as an example to weak and mortal men, so that they would not depart from justice on account of wrongs or insults or pain or death, if they did not know that He Himself experienced all these things?

In order to inspire others to trust and follow Him, Anselm argues, the God-Man must make His concern for others public and credible. Shared experiences are often the basis for deep and affectionate loyalties (which is one reason politicians and suitors so frequently appeal to them). Anselm’s God-Man, both a politician and a suitor, goes even further: He not only shares the experience, but shares willingly the suffering that others undergo unwillingly, earning not just the affection born in solidarity but the admiration and gratitude reserved for the magnanimous. The God-Man\textsuperscript{64}

freely offered to the Father what He was never going to lose as a result of any necessity; and He paid on behalf of sinners that which He did not already owe for Himself. Therefore, it is much more the case that He gave an example, in order that no single human being would hesitate (when reason demands it) to render to God on behalf of himself that [life] which one day he will summarily lose. For although He did not at all need to do so for Himself, and

\textsuperscript{63} CDH II.11 (H&R): Exemplum autem se ipsum quomodo daret infirmis et mortalibus, ut propter iniurias aut contumelias aut dolores aut mortem a iustititia non recederent, si ipsum haec omnia sentire non agnoscerent?

\textsuperscript{64} CDH II.18 (H&R, except the last line): A. Nullus umquam homo moriendo praeter illum deo dedit quod aliquando necessitate perditurus non erat, aut solvit quod non debet. Ille vero sponte patri obtulit quod nulla necessitate umquam amissurus erat, et solvit pro peccatoribus quod pro se non debet. Quapropter ille multo magis dedit exemplum, ut unusquiseque quod aliquando incunctanter amissurus est, pro se ipso reddere deo cum ratio postulat non dubitet, qui cum nullatenus aut pro se indigeret aut cogeretur pro aliis, quibus nihil nisi poenam debet, tam pretiosam vitam, immo se ipsum, tantam scilicet personam tanta voluntate dedit. B. Multum propinquas desiderio meo.
although He was not at all compelled to do so for others to whom He owed only punishment, He gave with such great willingness so precious a life—indeed, His own self—i.e., so great a person. [To which Boso responds] You are approaching very close to what I desire.

Those people who will imitate the God-Man, persevering despite the difficulties, will be His beneficiaries and will benefit from His sacrifice.\(^65\)

Anselm suggests that both prayer\(^66\) and suffering\(^67\) are necessary for the imitation of this God-Man. Prayer is necessary to indicate humility before God, since “it is not appropriate for man to deal with God in any way as equal with equal.”\(^68\) Prayer is an admission that justice comes only from God, and is an act of submission to Him. By accepting suffering, on the other hand, man purifies his intention, because he accepts what is an apparent disadvantage out of love of justice. In both cases, the imitation of the God-Man’s example leads the imitator to act from justice rather than selfishness. Of

\(^{65}\) CDH II.19.

\(^{66}\) Cf. Anselm’s implicit response in CDH II.19 to Boso’s question (in I.19) about why we pray for forgiveness. Anselm says that the approach to God involves doing what Holy Scripture teaches, which includes praying the Lord’s Prayer. See also McIntyre’s treatment of this (\textit{St. Anselm and His Critics}, pp. 104-106), where he defends Anselm from critics who think he dispenses with the need for prayer. See also CDH I.20.

\(^{67}\) Cf. DCD 21 (H&R): “Because of the sin of our first parent human nature was made capable of suffering countless troubles; from this passibility grace works incorruptibility in us in many ways.” “Hominum enim natura propter peccatum primi parentis iam innumerabilium incommodorum facta est passibilis; ex qua passibilitate multis modis gratia nobis operatur incorruptibilitatem.” In CDH I. 20, Anselm agrees that the asceticism of a monk’s life does contribute to his salvation.

\(^{68}\) CDH. I.19.
course, these are not efficacious on their own,\textsuperscript{69} but when done for love of the God-Man, they serve to strengthen the bonds of solidarity.

But what about those who do not follow the example of the God-Man? Anselm by no means thinks that divine mercy empties hell (he denies that fallen angels can be saved, e.g.), which means that while God grants mercy to some, all remain governed according to the demands of maintaining perfect order. Those who choose not to follow Christ are free to do so, but if they reject the opportunity to accept satisfaction, it remains necessary to subdue them through punishment, out of a need to defend the right ordering of the universe.

There is more to the story of how God treats sinners. In the context of God’s strict justice, presumption is a sort of usurpation, an attempt to wrest control over some part of the universe from its rightful governor. If this analogy is apt, one might expect God to respond as even modern countries respond to treason—by neutralizing the traitor forever. The U. S. Constitution provides for execution in the case of treason, which obviously neutralizes the traitor from doing further harm to the state. The analogous divine punishment would be annihilation, especially if we allow with Anselm that the very existence of evil rational natures can still disrupt the just order. Anselm suggests that God’s decision not to annihilate demons and devils (“reduce them to nothing”\textsuperscript{70}) shows His mercy. Disobedient creatures deserve to lose not just eternal happiness but their existence itself, since in rejecting the purpose for which they were made—viz., to

\textsuperscript{69} CDH I.20.

\textsuperscript{70} CDH II. 16: “in nihilum redactus esset.”
preserve justice and obey God—they reject their very creation. For God not to do this is better than they deserve (*contra meritum*), presumably since even eternal punishment is still a better fate than not existing.

Given this framework, one might make the following conjecture: Anselm argues that the first sin was Satan’s attempt to wrest control over his own fate from God. The devil in sinning had to believe that it was possible for him to improve his fortunes by disobeying God and marring the perfectly ordered universe that had just been created. Adam’s sin was similar, in that he chose his own advantage over obedience to God. In the case of human sin, though, God’s response is rather interesting: rather than reducing Adam to nothing, or damning him to perdition straightaway, God instead increases the disorder in the universe by relaxing the tight control He formerly exercised. It’s as if He calls Adam’s bluff, telling him, “if you think you know how things ought to be, you are wrong, and I will make this clear to you by letting you try to rule this cosmos without me.” The suffering that is simply part of human condition would then be the consequence of God trying to communicate to human beings their absolute inability to rule their world. In this way, the punishment for original sin would be to grant human beings exactly what they think they want, the opportunity to rule, so that they can learn through abundant experience that they lack the power to discharge so great a responsibility.

7.5 Human Mercy, Political Roles, and the Responsibility for Right Order

Let us try to derive lessons for human governance from our discussion of divine governance. The Anselmian picture assumes that the ruler will be merciful, that she wants to restore her charges to justice, but that her ability to effect mercy will depend on
a variety of factors, some of which might be beyond her control. *Justice*, conceived as the preservation of right order, will reduce a leader’s flexibility to be patient and understanding, if the object of such merciful attitudes has the potential for creating more disorder. There is a sense in which justice is simply mercy shown to everyone—if you were alone, I could be patient with your unjust behavior, but because your behavior harms others for whom I am also responsible, I cannot continue to let you behave this way. Angelo’s observation in *Measure for Measure* nicely captures this relationship between justice and mercy: “I show [pity] most of all when I show justice, for then I pity those I do not know…”

Conceiving of justice in terms of right ordering, though, brings our attention to an often-overlooked topic: the way in which roles and duties limit our responsibilities. If (a) we are merciful, and (b) we are responsible for maintaining order in some area, and (c) our position of responsibility puts limits on our ability to effect the mercy we desire, then it behooves us to be very clear about the scope of our responsibility, so that we know exactly what has to be ordered and how much flexibility we have to effect mercy. Furthermore, if we know the scope of our responsibilities, we also have a better idea of the rights and powers we need to carry them out. Our distinction between the disposition of *being merciful* and the act of *effecting mercy* highlights the need for the *power* to make the benign disposition actual. A *right*, on this view, is a claim about the sphere of one’s

71 *Measure for Measure*, II.2.100-101. We have similar debates today about judges who show mercy to the criminals standing before them without considering the wider effects of such mercy. Proponents of strict sentencing guidelines argue that these are necessary to protect society against the tendency of judges to focus on the person in front of them rather than on the good of society as a whole.
authority: that in the just ordering of the cosmos, the person with the right has a responsibility that cannot be discharged without a certain power to act, and others are not to take this power away or interfere with it, on pain of disrupting the just ordering of the cosmos. We might say that a privilege is a similar protection of a power to act to discharge a responsibility, when the responsibility is not natural but human in origin.\(^7\)

These considerations suggest that the axioms of responsibility we defined in chapter two, showing the connections between power and responsibility, will be very important to the Anselmian picture of governing. Recall, the two axioms were characterized as follows:

1. **With power comes responsibility.** If one has the power to perform some action, then whether or not to carry out that action necessarily becomes an object of moral consideration.
2. **With responsibility comes the need for power.** If one finds oneself with responsibility of some sort, then one is also responsible for acquiring and maintaining the ability to discharge it.

Indeed, something like the first axiom can even be found in the rule of St. Benedict’s warning to the abbot: “To whom more has been given, from him more will be demanded.”\(^7\) Anselm seems to have adopted this axiom as a general principle of government—even divine government. The Benedictine formulation has to be changed slightly to apply to God within the scheme of *Cur Deus Homo*, since nothing is “given” to God, but our formulation from Chapter Two seems

\(^7\) In Anselm’s correspondence, a privilege seems almost always to have this meaning; indeed, it usually refers to a power granted by the Pope. See Aep 126, 149, 222, 303-4, 307, 454.

apt: with power comes responsibility. God has the power to create and perfectly govern the world, and so has the responsibility to order it perfectly. (As we suggested, this responsibility seems so intuitive that the seeming disorder of the world gives rise to theodicy questions.) The second axiom also seems applicable to God: with responsibility comes the need for the power to discharge it. This second axiom seems to be behind Anselm’s view that it is necessary for God to punish sinners. If we accept Anselm’s view that God’s punishing sinners is a defensive act against would-be usurpers, and that leaving usurpers unpunished will encourage others to rebel against justice, it follows that for God to be lenient with sinners would be a failure to discharge His responsibility for order (which it is impossible to attribute to God).

There are some features of this view that are particular to God. Only God has responsibility for all creation, and so only He is expected to govern everything perfectly. Only He has the power to govern everything perfectly, while still effecting mercy; a being with less power would not be able to preserve order except through punishing the usurpers. The obligations that creatures have to God are uniquely complete—there is nothing a creature possesses that it does not have from God, and so there is nothing it can give to God that it is not already obliged to give. And therefore the gravity of a creature’s offence against God is also unique—not only is it foolish, since implicit in any challenge against the rule of the universal governor is the desire to rule in His place, which is beyond the power of any creature, but it also shows complete ingratitude.

Anselm never wrote a political treatise, but we have significant biographical evidence that suggests Anselm believed the account of divine governance to be a good
model for human governance as well. We have already argued that the Benedictine Rule provides the basic outlines of Anselm’s view of good governance, and there are definite similarities between the Benedictine ideal of the abbot and Anselm’s picture of God as a governor of an honor. Most importantly, both views hold that the serious obligations of the person in authority require him to balance the goods of order with the need for mercy, and both have a forward-looking, results-oriented approach to the salvation of their charges. The goal of both is to get those under their authority to preserve justice for justice’s sake, and they are empowered to use any means that respects the freedom of each and that preserves just order for all.

Over the last fifteen years we have had a revolution in our historical picture of Anselm’s time of Archbishop of Canterbury. In 1987, Sally Vaughn published the first political biography of Anselm, which really shook up the established picture of his personality. Before Vaughn’s careful reconstruction of Anselm’s political activity as Archbishop, Anselm was thought to have been a saintly monk and brilliant thinker who was basically drafted out of the monastery and into service as archbishop, a task for which he found himself ill-suited and in over his head. Vaughn’s book argued that, on the contrary, Anselm seems to have sought the office of archbishop, to have had a very shrewd understanding of both William II and Henry I, and at the end of his life achieved or at least substantially advanced all his major political goals. Since Vaughn’s book,


Walter Fröhlich⁷⁶ and C. Warren Hollister⁷⁷ have agreed with Vaughn’s finding and have substantially fleshed out our understanding of this aspect of Anselm’s life. This revisionist scholarship provoked Sir Richard Southern, the author and primary champion of the view of Anselm as unfit for public life, into writing his elegant biography *Saint Anselm: A Portrait in a Landscape*. Southern views Anselm as a monastic intellectual concerned about nothing except the interior life and doing God’s will, and tries to give a different reading of Anselm’s archiepiscopal activity, showing him to be either a political naïf or politically lucky.⁷⁸ Southern’s view has not carried the day, though, in part because so much of it was built on his conviction, based on an act of historical imagination, that saintly monks cannot act effectively in a secular world.⁷⁹ As new evidence from the revisionists keeps suggesting Anselm’s political acumen, that conviction seems less tenable.

There still remains no serious historical study, in any language, of Anselm’s 475 surviving letters, and so for many of them we know very little about who were the recipients and what were the circumstances in which they were composed. But Vaughn, 

---


⁷⁸ Cf. Southern, *Portrait*, p. 439: “The work of administration was always a burden to him. He did not do it well because he had no instinctive sense of what was possible in earthly affairs….The thoughts and aims of worldly men puzzled him without arousing his interest….His [policies] had no clear aim, and whatever results they had were accidental. It is mistaken to measure the results against some large aim and to ask how far this aim was achieved.”

in her recent study of Anselm’s letters to women, argues that he seems to have reflected at some length about various political roles—including, perhaps most prominently, the roles of king and queen. Anselm’s letters often paint for his correspondent a picture of how to fulfill his or her role perfectly, whether that role be as a mother, wife, widow, monk, abbot, bishop, queen, or king. If this view is correct, it provides another point of contact with the political view advanced in *Cur Deus Homo*.

Throughout his archiepiscopacy, Anselm tried to fight for his vision of how his role and authority as Archbishop should be defined. Partly this was because of the confusion about the role of the Archbishop of Canterbury in the years just after the Norman Conquest, when the integration of the Anglo-Saxon kingdom and the Norman duchy was still in its tumultuous early stages. But mostly, Anselm placed such importance on the definition of his role because on his theory of government, a good and dutiful governor is judged by how well he keeps order in the sphere over which he has authority. On Anselm’s view, the King of England and Archbishop of Canterbury should work together as co-rulers of England, while the pope should see the primate of England as the ‘Pope of Another World’ (i.e., let him exercise his primacy without interference from Rome). Anselm eventually won the sought-for privileges from Pope Paschal II.

---


81 See Vaughn, *Anselm of Bec*, ch. 5, and idem., *Handmaidens*, pp. 163-4, for Anselm’s views about the relationship between king and archbishop, and for his partisans’ claims that the Pope had recognized Anselm as an equal.

82 Cf. Aep 303-4.
and he eventually, after two periods of exile from England, hammered out a working relationship with King Henry I that more or less satisfied his goals. As Vaughn argues, Anselm saw his job as archbishop to lead England *ad ordinem debitum*, “towards due order.” In a letter to the bishops of Ireland, Anselm described what he thought were his duties to Christ, the Church in England, and to his office:

I wished to cut back vices by pastoral rule, to subdue usurpers (*praesumptores*) and to call back to due order whatever was inordinate in that which is entrusted to me.

Vaughn argues forcefully that this statement of Anselm’s summarizes and explains the entire rationale for his political activities as archbishop:

In Anselm’s mind, nothing…took precedence over his God-given obligation to rule Canterbury and the churches of all Britain as a model archbishop and primate. Perhaps Canterbury’s destiny was not the most important issue in Christendom, but because God’s power had brought Anselm to the archbishopric and committed it to his care, Canterbury, like Bec beforehand, became his primary responsibility.

83 Cf. Aep 461-2, where Henry grants, and Anselm accepts, the *temporal* rule of all of England in Henry’s absence! This was a magnanimous gesture on Henry’s part, as a way of demonstrating to all that he and his archbishop were reconciled. Anselm was very old, and died a few months later, but it seems that the gesture was appreciated by everyone. See Hollister, op. cit., pp. 379-383, for evidence that Henry was treated with much sympathy by Anselm’s partisans.

84 Aep. 198. Cf. Vaughn, “Monastic Sources,” p. 53: “It is the thesis of this paper that Anselm’s conception of ‘due order’ was in fact his political philosophy…and that he saw as his duty the enforcement of this political philosophy—‘due order’—in a particular way.” See also idem., *Anselm of Bec*, ch. 5.

85 Aep. 198 (my translation): Proinde infulatus sedule quid Christo, quid eius ecclesiae pro loco, pro officio deberem cogitare coeipi, et pastorali regimen vitae resecare, praesumptores coercere, et quaeque inordinata, ut mea intererat, ad ordinem debitum volui revocare.

86 Vaughn, *Anselm of Bec*, p. 149.
Anselm believed he was required to devote all his energies to bringing order to that part of Christendom for which he was responsible. In his mind, God had put him in a position where his salvation required that he carry out his obligations perfectly. When he lost hope of being able to discharge this duty because the king denied him the necessary powers, he tried to resign, figuring that if God permitted that, He would also forgive him.87

We don’t have space to delve more deeply into Anselm’s political activities as archbishop, and so we will have to rely on this assessment by his leading living biographer to suggest that Anselm’s ideas about the justice of God influenced his ideas about politics of men. After our reconstruction of Anselm’s notions of justice as rectitude and mercy as restoration of rectitude within a regime of self-defense, we are able to glimpse a heretofore-unnoticed unity between Anselm’s intellectual and political lives, or more precisely between the historian’s Anselm and the philosopher’s Anselm. This brief survey of the most recent work by historians of the early Anglo-Norman state lends unexpected support to our reading of Anselm as a thinker whose ideas about divine justice and mercy are meant to have implications for human politics.

In Part Three, we will show how we can use these concepts to develop a liberalism that is an answer to John Rawls’s question: how is it possible for there to exist over time a just and stable society of citizens who are profoundly divided by religious, philosophical, and moral doctrines?88

87 Aep 193

88 Cf. PL p.4. This question is not exactly Rawls’s question, for the reasons developed in Part One.
PART THREE

INTRODUCTION TO PART THREE:

TOWARDS AN ANSELMIAN LIBERALISM

In Part Two, we developed in great detail Anselm’s theories about justice, mercy, right order, and good governance. Anselm’s most developed views are about the governance of God, but we argued that since they are cast in such abstract terms, and since all the central concepts are pure perfections and thus are applicable to creatures as well as to the Creator, that the account is not restricted to the case of God. In chapter seven, we also discussed some recent historical scholarship that suggests Anselm applied the very same principles of justice, mercy, rectitude, and good governance to his role as Archbishop of Canterbury. This gives us even more confidence to think that we can apply Anselm’s ideas to human politics.

At the end of Part One, however, we argued that in modern politics there was a need for a theory of mercy consistent with what we called ‘benign heteronomy.’ Even if Anselm’s views can apply to human politics, it is not clear that they can apply to the circumstances of modern pluralistic democracies. Chapter eight tries to show how this can be done. The trickiest question might be how to relate Anselm’s views about rectitude to the conditions of modern pluralism, since one obvious way of conceiving of
rectitude in politics—one leader trying to impose his own view of order (or his conception of God’s view of order) on society—is unworkable given what Rawls calls the ‘fact of pluralism.’ So part of the task of chapter eight will be to show how to reconcile these Anselmian principles of governing for order with the conditions of pluralism and democracy.

Chapter nine fleshes out the theory sketched in chapter eight by developing an Anselmian account of criminal punishment. This addresses what we argued was a deep problem for other liberal theories—that they lacked a sound account of criminal punishment consistent with the principles of their liberalism. This problem opens up Rawls’s theory and similar theories of contractarian liberalism to the charge of utopianism. In chapter nine, we compare the Anselmian theory of punishment to the leading theories of punishment in the literature, and show how it shares their strengths and improves upon some of their weaknesses. At the end of chapter nine, we hope, the reader will have a fuller picture of this Anselmian liberalism of mercy.
CHAPTER EIGHT

THE LIBERALISM OF MERCY

8.1 Introduction

Anselm is one of very few brilliant philosophers to have also been an important political ruler. Marcus Aurelius is a parallel in antiquity, while there might not be an example in modernity. Since the Stoicism of Aurelius does not seem to have formed a blueprint for how he governed Rome (nor is it clear how it could have), Anselm might be the only ruler to have developed a plausible philosophy of justice and have attempted to use it to rule. Plato might not have had an abbot/archbishop in mind when he dreamed of the philosopher king, but it seems that Anselm might have been one of its only genuine instances.

In chapter seven we suggested that current work in the history of the Anglo-Norman state is helping to piece together the narrative of Anselm’s reign as Archbishop of Canterbury. It remains to be seen to what degree Anselm’s speculative works can be integrated into the that story, though I have suggested that there is a tighter fit than historians have previously supposed. This work of integration is made quite tricky, though, by a contingent fact of history: although Anselm was elected Archbishop in 1093 and occupied the office until his death in 1109, Anselm spent two three year periods, 1097-1100 and 1103-1106, in exile over disputes with the king. As one might suspect,
relations with the throne were also far from cordial in those years when Anselm was not in exile. For over nine years, Anselm sought royal permission to hold a council where he could pass legislation for the English church; only in 1102—after the death of William II, the ascension of Henry I to power, and Henry’s consolidation of that power after a war with his older brother, Robert of Normandy—did Anselm receive this permission. Anselm spent most of his tenure seeking the power to discharge the duties that accompanied his role. It was endlessly frustrating for him to be denied this power, or to be offered it only in exchange for compromising some other aspect of his office. It is also frustrating from our perspective, for we are not given the chance to see how his views of justice, punishment, mercy, and order worked together when put into effect. As we saw earlier, Anselm in a letter to the bishops of Ireland describes his policies in terms that evoke his speculative writings:  

1 Aep. 198.: Proinde infulatus sedule quid Christo, quid eius ecclesiae pro loco, pro officio deberem cogitare coepi, et pastorali regimine vitia resecare, praesumptores coercere, et quaeque inordinata, ut mea intererat, ad ordinem debitum volui revocare.

I wished to cut back evils by pastoral rule, to restrain usurpers (praesumptores) and to call back to due order whatever was inordinate in that which is entrusted to me.

Here the motifs of responsibility, order, punishment, and the restraint of the presumptuous are all present.

It is not our purpose here to show the degree to which even within the constraints of the politics of the time, his policies did enact his philosophical ideals.  

2 I do believe, though, that there are deep consistencies between his philosophical ideas and his actual behavior.
principles, which of necessity will go beyond Anselm’s actual policies. It will be helpful here to lay out those principles to see just where Anselm’s own ideas end and the Anselmian development begins.

8.2 Anselmian principles of political justice

Although this division is somewhat arbitrary, we can discern in Anselm’s letters and speculative writings nine principles regarding the idea of due order and the leader’s responsibility for this, and eleven principles regarding the importance of stability for order and how punishment (and mercy) increase stability. All these principles are found in Anselm’s writings, though not only there—many of them can be found in Rawls or in other authors, or might seem simply to be common sense. Indeed, we introduced versions of some of these principles in chapter two. After laying these out, we will need to add some further principles so as to apply them to liberal democracies.

**Anselm’s principles regarding responsibilities and due order:**

1. *Powers*\(^3\) *create responsibilities, which create roles with rights and duties.* Anselm argues that the evident *ability* of man to know and love the good makes manifest his *duty* to know and love the good, and the highest Good in particular (M 68, CDH II.1). God’s ability to rule perfectly requires Him to rule perfectly (CDH I.15). God’s decision to create requires Him to act with justice towards creatures (M 69, 70, 74, CDH II.5). Man violates his duties whenever he fails to exercise his intellectual

\(^3\) Obviously, if there are powers not subject to voluntary control (the power of digestion, or of growing hair, e.g.), these do not create responsibilities. This claim only applies to voluntary or rational powers. I thank Stephen Dumont for this distinction.
powers responsibly by knowing and loving God (CDH I.11, M 71). If ever a man were to keep this duty, he would have a right (it would be unjust were he not) to have his love requited (M 70, CDH II.2). Insofar as it is possible for man to understand something of God in this life, he ought to do so (CDH I.1).

2. **Roles can sometimes be given to us by someone else or by an institution.** Duties *always come with the role; powers often do as well.* This is a basic principle of the rule of St. Benedict, which Anselm applies over and again. He even claims in several places that he was pressed into his role as author under obedience to his brothers (M proem., CDH proem., DIV proem.). Together with #1, we can divide roles into *natural* and *acquired*, depending on whether the role is a result of the need to regulate a natural power according to morality, or whether someone else bestows the role on us.

3. **To step out of one’s role is presumption, and a source of disorder.** Anselm describes all sin as the sin of presumption in DIV 1. In DIV 11 he contrasts the magnanimity of Christ in “emptying himself” (cf. Phil. 2:7) with the presumption of the devil and the sinner who tries to seize equality with God. In CDH I.7, and again in I. 20, Anselm claims that one has to have the proper jurisdiction even to do something right and just, or else it is presumptuous or unjust.

4. **Sometimes the role comes not with power but with a duty to acquire power.** Anselm felt that his role as Archbishop impressed upon him the duty to call primatial councils so that he could govern, and he spend the first nine years of his reign seeking this power from the king.
5. All people are responsible for themselves at some level. This responsibility is inalienable. According to Benedict’s Rule (ch. 2), even monks cannot entirely give up their responsibility for their souls to the abbot, because they retain free will. Anselm’s doctrine of the will prohibits even God from taking away free choice (DLA 8). Given #1, it follows that because a person has the power to choose to love God out of justice, then he has the responsibility to choose to do so (DLA 2, DCD 27).

6. In a rightly-ordered society, roles are not redundant, so if one has a responsibility, it is important that it be discharged. Anselm rebuffed attempts by others to infringe on his role, whether they were attempts by the Pope (Aep 214, 222) or the other bishops in England (Aep 198, 303, 442-5) or the king (Aep 176). Anselm saw his role as something either to be taken up with all its responsibilities or relinquished, but the responsibilities were forever tied to the role. When he despaired of discharging his responsibilities, he asked the Pope to relieve him of the office (Aep 193), but he fought for years to make sure that the powers and responsibilities of the office not be diminished.

7. If I am responsible for others, I am responsible for those under me, and to those not under me. I answer for my charges to others. I am a mediator. This is clearest in the case of the Deus-homo, who is the mediator answering to God for man. Anselm thought of his own job as bishop as answering to God (and his vicar, the Pope) for the order of his church—hence his anguish at not being able to order it properly (Aep 193). Anselm would blame an abbot for the misbehavior of his monks, and blame a bishop for the misbehavior of an abbot (Aep 195) or of his priests (Aep 172). Of course, one can have responsibilities to the same person for whom one is
responsible—Anselm was responsible for the soul of the king and responsible to the king for the proper governance of Church property.

8. *I'm accountable for good results, not just for doing my best. It is possible to do all that is within my power, and yet to fail in my responsibility.* The office of abbot or bishop is a burden and a danger to one’s soul, which one ought not desire (Aep 61). Anselm felt the burden of failure, despite his best efforts: “I grieve that I am a bishop because, with my sins impeding me, I do not carry out the office of a bishop….I collapse under the burden because, more than seems believable, I suffer from lack of strength, virtue, diligence, and knowledge adequate to so great an office.” Weakness is not an excuse for disorder—hence Anselm’s response to Boso that the mere inability of man to restore God’s honor doesn’t release man from the burden of restoring that honor (CDH I.24). This explains the duty to acquire power in #4. On the other hand, nobody is responsible for the will of another (see # 5-6).

9. *Having responsibility requires that I achieve right order in the area for which I am responsible.* Leaders provide order—this is tautological, because leadership is necessarily teleological, and to lead toward an end is to order. We have already discussed at length how Anselm conceived of his own leadership and God’s as instilling order.

---

4 Aep. 193 (WF translation): *Doleo me esse episcopum, quia peccatis meis facientibus non ago episcopi officium….Oneri quidem succumbo, quia virium, virtutum, industriae, scientiae tanto officio competentium inopiam plus quam credibile videatur patior.*
Anselm’s principles having to do with stability, punishment, and order:

10. Responsibility requires a concern for stability, namely that the order be as enduring as possible given the nature of the responsibility. (This is obviously a point of contact with Rawls.) Anselm tried to preserve the rights of Canterbury against encroachment from all sides, in part because he felt he had an obligation not to lose any of what God had entrusted to him. As we have already seen, Anselm insists that God’s governance be in perfect order, and everything that ever exists will always be under God’s governance (CDH I.15).

11. Stable rule requires that people buy into the principles of the rule, or at minimum have a principled affection and trust for the leader. (This is another point of contact with Rawls.) Anselm defines justice as rectitude of the will, and later he describes rectitude as an honor, with all the relationships of loyalty and obedience that concept implies. The idea of an honor helps answer a worry that one might have with the DV 12 definition of justice: it would seem that if someone does not know what is God’s will, he could not be just; yet creatures are never going to be able to know God’s will, since that is identical with the divine essence. But if loyalty to a divine mediator is sufficient to restore God’s honor, then one needn’t know God’s will directly, but need only be a member of the mediator’s party. It is possible to be loyal without being wise, just as it is possible for one person to love another without understanding the other completely. So long as that loyalty is a stable love for the mediator, and the mediator’s will is just, such loyalty should count as justice.

5 It should be clear that several of these principles have much in common with what Rawls has to say about stability.
12. *One cannot rule stably simply by appealing to the self-interest of the ruled.* (This is analogous to Rawls’s criticism of the modus vivendi.) Anselm claims in DC III.13 that God gave people justice for His own glory, and he argues in DCD 13 (following DV 12) that one cannot pursue justice out of one’s own self-interested advantage. God cannot be glorified, i.e., He cannot receive the honor due Him, so long as people lack a principled affection for His will.

13. *Punishing the presumptuous can help restore order.* The act of punishing can put a presumptuous person in his place and make it clear that he is subject to the punisher. Since presumption takes away the deference and obedience and loyalty that is due another, the punisher can regain some of what he lost by forcing the presumptuous person to submit to his will (CDH I.14).

14. *The duty to preserve order might compel a leader to punish.* Since all rational beings have free will, it is impossible for a rational being’s love to be coerced (see #5-6, 8). Therefore punishment cannot force someone to have a just will. But punishment can result in the submission of the person in all matters of behavior (#13), and it can even provoke a certain amount of cooperation (at least out of self-interested fear). If the goals of the leader are just, then this submission and limited cooperation will further the ends of justice to some degree, and will prevent the presumptuous person from creating further disorder. In the case of a particular recalcitrant person with free will, this level of order is the most that can be asked of a leader (Aep 195). If it is possible for him to restore this much order, and not punishing would lead to more disorder, then the leader has a duty to punish (CDH I.15).
15. There is a point at which too much punishment detracts from the order for which a leader is responsible. While in each case taken individually, the leader cannot be faulted for the recalcitrant disobedience of a free rational being, he can be faulted if he does not command respect and affection from enough people. This is the argument Anselm gives for why it would be wrong for God not to save some human beings (I.25, II.4-5): Part of the right order of creation is that a certain number of souls be in heaven, and while each particular decision to punish a sinful creature is clearly just, it would be a sign of poor governance were no creature to be saved. God should not allow the weakness of creatures to mar the beauty appropriate to heaven.

16. Mercy leads to affection for the merciful person, because it is a sign of that person’s goodness. Anselm’s discussion of justice and mercy in P 8-11 was confused in part because it tried to claim that mercy was a pure perfection in God. By the CDH, it was clear to Anselm that mercy cannot be a pure perfection in God, since mercy necessarily requires a relation to someone who lacks a just will, and since God was free to create the world or not (CDH II.5), a necessary relation to the world couldn’t be part of His essence. What mercy does show is a ‘disposition’ to be merciful in conditions where mercy is possible, and that disposition (which in God is identical with His nature) could plausibly be described as goodness or love. Showing mercy reveals something of the character or nature of the one doing the showing; as Anselm writes: 6 “You are merciful, then, because you are fully and supremely good,” and

---

again,⁷ “God, saving us, has shown the greatness of his love and devotion towards us.” Since rational beings tend to love the good when they apprehend it, showing them mercy leads to their affection. On the other hand, mercy shown out of weakness does not yield this affection, but rather scorn for the weakness (CDH I.24).

17. If punishment is necessary, the leader should punish reluctantly and not out of a joy of punishing. Anselm says that the love of justice ought to draw one to punish reluctantly, but that to punish eagerly is characteristic of the devil’s malice (CDH I.7). Showing mercy whenever possible reveals this reluctance to punish, and demonstrates the good will of the leaders (Aep 97, 104, 137).

18. Ideally, the punishment will lead the punished to have a conversion, so that he develops a principled affection for the foundation of the regime. Anselm has a letter (Aep 233) explaining why it is better for a monk to be beaten publicly at the command of his superior than it is for him to scourge himself or to be beaten in private. The public beating at the command of the superior gives the monk the chance to submit his will to the obedience of another, thus giving him a chance to grow in the virtue of obedience as well as in humility, while in the other forms of beating and scourging one retains command of himself and does not subject himself before man. As we saw, Anselm believes that the “severity of the Rule” (Aep 97) is to be applied only reluctantly.

19. To show mercy responsibly requires that one be able to maintain order while taking the risk that mercy involves. Therefore, only the sufficiently powerful can show mercy.

⁷ CDH I.3: “…restituit, tanto maiorem dilectionem erga nos et pietatem monstravit.”
This principle, which is based on several of the foregoing (# 4,9,10,16) is a key assumption in Anselm’s argument that only the God-Man is able to effect the salvation of any human beings (CDH I.25).

20. Even in contexts where punishment isn’t relevant, a leader still needs to bring others to a principled support of right order in the relevant sphere. Therefore, he needs to be merciful, realistic, solicitous, proselytizing, patient, and appropriately demanding (even infinitely demanding if there is a strong enough idea of order). Anselm’s letters show that he lived up to his own injunction that one should punish only reluctantly. Much of his correspondence consists of exhortations to his correspondents, whether lay or religious or clerical, encouraging them to live their vocation as perfectly as they can.\(^8\)

8.3 Can There Be Contemporary Applications of Monastic Principles?

In Book V of the *Nichomachean Ethics*, Aristotle notes that he is describing the justice of citizens and not the justice of a master towards his slave or a father towards his children.\(^9\) He even suggests that relations of this latter sort might not even be governed by justice at all, since the slave and the child are extensions of the master and the father, respectively, and one does not show justice to oneself. It should be evident that the principles of governance and order that we find in Anselm are not primarily about justice between equal citizens, but are instead more like the justice of the master or father.

\(^8\) This is the thesis of Sally N. Vaughn’s study of Anselm’s letters to women, *St. Anselm and the Handmaidens of God* (Brepols Publishers, 2002).

Unlike Aristotle, though, Anselm does not regard the inferiors as mere extensions of the superior, but instead he regards them as independent moral agents with freedom and innate dignity— in the striking image from the Cur Deus Homo, they are valuable pearls, even if through sin they find themselves covered in mire.¹⁰

Most liberal theorists follow Aristotle’s lead and restrict their discussion of justice to relations between equals. In Part I, however, we saw that there are some deep problems with this restriction—it imposes social forms which assume that all are equal on people who aren’t equal or who don’t regard themselves as equal, and it creates significant impediments to thinking seriously about the moral issues of punishment. I want to suggest that rather than trying to force democratic politics into an ill-fitting egalitarianism, we can more profitably conceive of modern liberal democracies as very intricate networks of overlapping hierarchical jurisdictions combined with a set of institutional and moral checks against destabilizing accumulations of power. To put it perhaps too paradoxically, the idea is to combine the Rule of St. Benedict with Shklar’s liberalism of fear. The key to applying these monastic principles to liberal democracies is the following, what we might call The Principle of Monastic Liberalism:

In democracies, all citizens can be thought to have the role of a ruler or governor with real but very circumscribed powers, and thus real though diffuse responsibilities.

If we emphasize this aspect of democratic citizenship—the citizen in her role as governor—then Anselm’s principles can start to look rather applicable.

¹⁰ CDH I.19.
In a democracy, almost everyone is given the power to vote, and therefore, if we apply Anselmian principle #1, the responsibility to vote well. If I have the power to associate and debate about politics, I’m responsible for how I use that power. If I have the power to run for office, I’m responsible for deciding whether or not to run. I’m not just responsible for acting honorably, I’m also responsible for the outcome of my actions—if there’s more I could have done to prevent a demagogue or an otherwise dangerous person from winning, then I’m responsible in a diffuse way for the bad result (cf. principle # 8).

Responsible to whom? Other citizens, future generations, other countries, even—to those who can’t vote, and to those affected by political action. (At this point we might hear echoes of the warning attributed to Edmund Burke, “All that is necessary for the triumph of evil is that good men do nothing.”)

On this view, democratic political equality is the result of the equality of the power given to each citizen to cast only one vote. Equality of this sort is the defining feature of democratic politics; to have it is to have a democracy, not to have it is not to have a democracy. (Note that not all persons under a regime need be citizens in this sense.) But this narrow sort of equality is entirely consistent with all sorts of political inequalities in other areas. The rich, if they have a greater ability to affect politics, have a greater responsibility for the outcome. The same is true for the famous and the well-respected, the culturally influential, those who are leaders of various groups of civil society, the naturally charismatic, even the heads of families—to the degree that political influence is available to them, they are morally responsible for how they use that influence, and are morally responsible for the various political outcomes. When a person is a citizen, she has a role that comes with the power to engage in politics and the duty to
do that well (#2). Since citizens have different talents and different opportunities to bring their talents to bear on politics, their being citizens bestows on them not only the bare duty of voting well, but the duty of using all their talents to make society better ordered.\textsuperscript{11}

‘A rightly ordered society’ will mean different things to different people in the sort of pluralist societies that, as Rawls argues, liberal democratic political institutions are likely to produce.\textsuperscript{12} One person might think that a rightly ordered society permits abortion, while his neighbor might think that it does not. This is not to imply that at any given moment there is not a right answer to the question of how society should be ordered now. It might be the case that from a God’s eye point of view there is exactly one way that this society should be ordered right now, even all the way down to the hairs of a person’s head.\textsuperscript{13} But among those of us who do not have a God’s eye point of view, there is going to be disagreement, even deep disagreement, about how to order society. Since human knowledge is finite, our attempts to order society must take into account the agent’s partial ignorance about what is the actual right ordering of all creation at this moment.\textsuperscript{14} This ignorance will fall into at least three categories: ignorance as to 1) the

\begin{flushright}
\textsuperscript{11} Note that this observation does not yet disagree with Rawls’s assumption that one’s wealth, natural gifts, etc. are morally irrelevant to questions of justice and ought to be behind the veil of ignorance in the original position. The claim here is perfectly general: parties in something like the original position could all agree that if one has more talents, one has the moral responsibility to use them well.

\textsuperscript{12} Cf. PL pp. xxvi, 4, 36-8.

\textsuperscript{13} Cf. Mt. 10:29-30.

\textsuperscript{14} As we saw in chapter six, Anselm emphatically rejects that either God or His plan for creation are \textit{completely} unknowable to us. It seems that everyone, even the pragmatic skeptic, has some views about how things ought to be—the pragmatic skeptic thinks, first, that we ought not to act on information that is uncertain unless we are
nature of what it is good to do or even of the Good itself; 2) the human abilities of
particular people at any given moment; 3) the world-historical circumstances in which
particular people may act. I might not know the right thing to do in a given situation
because I don’t know what it is good to do, either because I lack certain virtues that help
me apprehend the good, or because the nature of what it is good to do is unknowable in
itself, or because I am not in the right epistemic circumstances to apprehend what it is
good to do. It might be that I have some thoughts that the world would be good (or at
least better) if it were ordered in such and such way, but I don’t know whether I can bring
it about, or whether anyone can bring it about. This problem is especially acute when our
decisions depend on how other people will behave—not just because of the theoretical
problem of other minds or the problems of predicting the responses of free rational beings
(both of which could be significant sources of ignorance), but also because of the
practical problems a leader has in assessing untapped potential in himself or another.
Finally, even were I to have a definite idea of how things ought to be and a good sense of
how people will react, I could not take every eventuality into account, since all sorts of
circumstances could affect my plans.

All these limitations ought to urge us to a practical modesty when developing our
ideas about how things ought to be and in trying to determine how best to bring about
that state of affairs. But if we have any power at all to affect how the world will be, then
we have a responsibility to determine how to use that power, which means we are forced
compelled to act, and second, that most information is uncertain. The first is a moral
claim about how some part of the world ought to be, and however modest such claims
are, they are sufficient to prevent one from claiming total ignorance of how things ought
to be.
to develop some views about what the world ought to be like and to concern ourselves
with the practical question of how to make the world correspond to this vision (principle
#9 above). At minimum, each rational person has a responsibility to determine her own
behavior and attitudes (#5). If Rawls is right about the nature of political stability, even
the seemingly private and personal question of whether or not one has a principled
affection for the regime might have important consequences for whether the regime
continues on. As we argued above, those who have a greater ability to influence the
culture have a greater duty to exercise their powers well. This in turn requires them to
develop a more extensive set of views about how things ought to be, and it places on
them a greater responsibility for how they turn out.

8.4 The Rights and Responsibilities of Citizens

In a democracy, every citizen with the power to vote has a corresponding
responsibility to determine to some extent how the society should be ordered. Citizens
can conceive of this duty in two ways, corresponding to two different theories of
democratic representation. At least, they have the duty of determining which people will
govern, and determining whether those who govern, govern well. On the other hand,
some might argue that the citizens have a more robust duty to determine in significant
detail how society should be ordered, so that they can pick representatives who will carry
out the details of their policy wishes. On the first view, citizens conceive of elected
representatives as leaders, while on the second view the representatives are closer to
employees of their constituents. These are extremes, and in most circumstances the
representative will have to be more independent than on the employee model (because he
will be faced with particular judgments for which his constituents have not provided clear guidance), yet less independent than on the first model (since the voters have a responsibility to develop some sort of standard by which to judge the performance of the representative).

In either case, democracies should be viewed not primarily in terms of the equal distribution of rights, but rather in terms of the equal distribution of the responsibility for determining which citizens will govern, and determining whether they govern well. On this view, the political rights that Americans habitually attach to their conceptions of democracy should primarily be seen as being in the service of the discharge of the citizens’ political responsibilities. Freedom of speech, the right of assembly, freedom of the press, and so on are necessary preconditions for democratic citizens to discharge their rights of responsible governance. Because they are in the service of bringing about right social order, such rights are not absolute, and do not exist when their exercise would conflict with right order. In principle, therefore, irresponsible uses of these freedoms can be outlawed. Of course, the liberalism of fear suggests that there are often good reasons why a society might wish to extend these freedoms more broadly than this political teleology might suggest. One might even propose, on liberalism of fear grounds, a right to be deliberately and consciously socially irresponsible—e.g., extending the protections of free speech to music that fosters racial disunity and anarchic rebellion against legitimate public authorities. But such rights are extended or not as a matter of prudence. They are not essential democratic rights because they are not essential to the discharge of basic democratic responsibilities.
While it would take us too far astray to derive a full-fledged theory of individual rights from the concept of due social order, we can provide a brief sketch of some characteristics of such a derivation. If we adopt a God’s eye perspective, a rightly ordered society would be a society in which all persons and things would exist in the correct relationship to each other and to the wider social order. Individual rights would then be the correct relationships the individual has to the larger order. If we don’t know the whole order, we don’t know all the rights an individual has. We can talk about *prima facie* rights in a provisional fashion—as a rule of thumb, everyone has a right to free speech, or right to property, etc. But such rights are not absolute, and do not exist when their exercise would conflict with the right order. A political leader cannot always speak freely, for example: the duties of office grant greater weight to injudicious comments, and so it can really be *wrong* to speak injudiciously. So when the exclamation of an exasperated Henry II—“Who will rid me of this meddlesome priest?”—leads to Beckett’s murder, we can hold Henry responsible for the murder, whether or not he intended for his courtiers to carry out the act. Similarly, when Ronald Reagan jokes during a 1984 microphone test, “My fellow Americans, I am pleased to tell you I just signed legislation which outlaws Russia forever. The bombing begins in five minutes,” we could argue that, were it not for the positive law of the First Amendment, he had no right to do so, because in the nervous context of the Cold War it was wrong to do so.

This view of politics makes rights secondary, and emphasizes instead the political goals of stability and social order. This is markedly different from standard accounts of liberalism, which make egalitarian rights theoretically central. It might thus be helpful to
recall the two deep reasons for this difference: the imperative of stability and the need for mercy.

As we argued in chapter two, a leader has a basic responsibility to maintain the stability of the regime (see also principle #10 above). In a democracy, in which all citizens are viewed as having responsibility for good governance and for social order, it follows that it is a basic responsibility of all citizens to promote the stability of the regime. If we agree with Rawls that a state should be thought to be of infinite duration, then the imperative of stability commits us to two further conditions: i) the principles of a stable regime cannot be based on contingent features of a particular people or time, such as the current balance of power between religious groups (see principle #12); ii) such principles must not assume that a society will be homogeneous. In a democracy, this means that all citizens have a responsibility to subscribe to and foster affection for political principles that balance the permanent features of the human condition with the heterogeneity of beliefs and cultures over which a stable regime must rule. These responsibilities are more basic than any other political and social responsibilities, since the continuation of the regime is a precondition for all other activities in the political and social order. If someone were to put the future existence or right order of the regime in serious danger by an act of speech or by the publication of a journal, then the obligation to promote stability would trump any right to freedom of speech or the press. This is true not just in an extreme “state of emergency,” where the threat to the social order is clear and present, but also in the case of a constant series of acts with the purpose of
undermining citizens’ loyalty to the regime over time. If we agree with Rawls and Anselm that stability consists not just in behavior that supports the regime, but also in a principled affection for the regime, then we are forced to conclude that each citizen ought to be concerned with the thoughts, attitudes, and opinions of the other citizens, and not just how they behave. Presumably, it would follow that the state has a legitimate role in fostering certain ideas and attitudes and discouraging others. In a liberalism of fear that role ought to be highly scrutinized. Nevertheless, it is a real and necessary role, which it would be irresponsible for citizens to oppose in the name of a false conception of neutrality or rights.

As we saw in Part One, Rawls believes that political justification rests in significant part upon the moral psychology of the citizens, and that the moral psychology of citizens is very much a concern of the civil authorities, pace Locke’s first Letter on Toleration. We suggested that this view explains why Rawls and other contemporary liberals believe that some sort of civic education in the principles of liberalism is necessary for the stability of the liberal regime. Rawls hopes that this education will be uncontroversial if all citizens have an overlapping consensus about the principles of political liberalism. This hope is unlikely to be vindicated, though, since the civic education will be at least as problematic for integralists as is the general scheme of Rawlsian liberalism, and for the same reasons. An integralist will be unlikely to accept Rawls’s suggestion that all children be educated about their “constitutional and civic

\[15\] Again, it is important that there be significant checks and balances, both institutional and moral, against the casual abrogation of important political freedoms in the name of transient state “necessities.”
rights,” especially “that apostasy is not a legal crime.”\textsuperscript{16} since someone who wishes to integrate her life around her comprehensive doctrine and even to bind her future self to that doctrine will not be likely to want her children to be encouraged to treat that doctrine as just one of many equally valid life choices they can make. This creates a problem for Rawls: the Rawlsian state has an interest in promoting tolerance and autonomy among its citizens, and especially to the young through civic education, but that civic education will be destabilizing if imposed upon the children of integralists. On the other hand, an Anselmian view would allow us to sidestep this issue. Such a view would agree with the above argument about the necessity of civic education for promoting stability for the right reasons across generations. But since the Anselmian view does not advocate autonomy or Rawlsian tolerance, integralists would not have the same problems with it.

If Anselmian civic education would not promote tolerance and autonomy, it would have to promote something similar enough to fill their place. On Rawls’s view, citizens who had internalized the virtues of tolerance and reasonableness would voluntarily abstain from using the coercive powers of society to advance those aspects of their comprehensive doctrines that could not meet the standards of public reason. By not imposing their views of the world on others unless they think the view is publicly defensible, such tolerant citizens promote stability by respecting the dignity of each citizen, or in Rawlsian language, by protecting the social bases of self-respect. Of course, this view assumes that the dignity of citizens is violated when the citizens are subject to the will of another (unless that will is one that is defensible from the vantage of the

\textsuperscript{16} JAF, §47.4, p. 146.
original position, in which case one is not subject to the arbitrary will of another, but is subject to a specification of one’s own rational will). On Anselm’s view, being subject to the will of another does not in itself cause the one subject to lose his dignity, because of the virtue of mercy.

The merciful person refuses to treat people as objects, but instead respects the depths of their personal freedom, and tries to win the person over to justice (#12). A merciful person can have a comprehensive doctrine that is not publicly accessible, and can try to advance it in the public forum, without thereby violating the social bases of self-respect or leading to instability. This is because the merciful person is not interested in advancing his view of due order by bribes, threats, or appeals to citizens’ self-interest, but rather wants his fellow citizens to accept his vision of due order as their own (#11). The merciful citizen’s primary methods for advancing his idea of social order are thus friendship, acts of solidarity, patient listening to the views of others, thoughtful study of a problem from all viewpoints, solicitude, working to understand the defects and limitations of others, and the development of sensitivity to the effects of any proposed action on the persons affected (#20). A merciful person will be reluctant to use coercion to bring about his view of the social order (#17), although he will be prepared to do so if it will promote the stable ordering of society (#13-15). If at some point his view cannot command enough support to move forward, then in the interests of stability he has to wait until he can acquire enough power (in the form of political support or influence) to be able to promote his view with minimal use of coercion while still retaining the flexibility to be merciful to those who are recalcitrant (#19). In the case of major changes to society which are likely to meet with widespread or deeply felt opposition, the merciful person
will move slowly so as to minimize instability and the need to punish, and will even be prepared to postpone his plans indefinitely if their imposition will lead to great enough short-term instability.

Although on this view civic education will not promote the values of tolerance and autonomy, they will promote mercy, a sense of responsibility for the common good, a respect for the freedom and dignity of the person, an aversion to manipulating people, and the value of the auxiliary virtues of patience, solicitude, and friendship, *inter alia*. Ideally, all citizens will internalize and accept this view of the good citizen. If so, then we have an answer to the question that animates Rawls’s *Political Liberalism*: How is it possible to have a stable liberal democracy in circumstances of religious, philosophical, and moral pluralism?
CHAPTER NINE

PUNISHING FOR STABILITY

9.1 Introduction

It is unlikely, of course, that all citizens will internalize and successfully fulfill all the demands of the view of the good citizen sketched in the previous chapter. We saw in chapter four that Rawlsian liberalism is unable to justify an attractive and viable system of criminal punishment. Can the Anselmian view do better? Any successful theory of punishment would have to be consistent with the imperative of stability, which in chapter two we described in the following way:

Leaders of a regime are responsible for conducting themselves according to principles that garner sufficient support so that the regime (i) is able to sustain itself in the face of external and internal attack, and (ii) is able to instill in future generations a respect for the regime and its principles.

Can the Anselmian view developed here propose a view of punishment that protects public order from internal attack, but does so in a way that instills an enduring respect for the regime and its principles over time, even among (at least some) of those punished (#19)? We have two reasons to think that it can: First, the Anselmian notion of rectitude solves the conceptual paradoxes of mercy and justice that make it so difficult to settle on a stable and humane corrections policy. Second, the Anselmian liberalism sketched here
is consistent with ‘restorative justice’ and ‘moral education’ approaches to punishment, which are the most humane accounts of punishment proposed in the literature. Let us examine these reasons in more depth.

9.2 The paradoxes of mercy and justice

1. Two Paradoxes

In the *Proslogion*, Anselm stumbled across two paradoxes that mercy creates for justice. Following Jeffrie Murphy, we can refer to them as the inconsistency paradox and the equal protection paradox.¹ The inconsistency paradox, taken from P 9, suggests that the concepts of mercy and justice are not consistent. Here’s how Murphy summarizes it:

If we simply use the term ‘mercy’ to refer to certain of the demands of justice (e.g., the demand for individuation), then mercy ceases to be an autonomous virtue and instead becomes a part of (is reducible to a part of) justice. It thus becomes obligatory, and all the talk about gifts, acts of grace, supererogation, and compassion becomes quite beside the point. If, on the other hand, mercy is totally different from justice and actually requires (or permits) that justice sometimes be set aside, it then counsels injustice. In short, mercy is either a vice (injustice) or redundant (a part of justice).

Martha Nussbaum, commenting on an ancient version of the same problem, points out that there is a difference between mercy, which mitigates justice, and equity, which applies general rules of justice to a particular case.² It is one thing to determine what is the appropriate punishment for this particular person who committed this particular crime.


under these specific circumstances, and something quite different to decide not to apply that punishment. If we distinguish between equity and mercy, then the problem can be framed as follows: *since equity is a part of justice, then mercy is either identical with equity, or it is unjust and equity is merciless*. As Anselm realizes, this is a serious problem for understanding God’s justice: since God is perfectly equitable, judging in every particular, it would seem that either He is merciful and not equitable, or (given the doctrine of original sin) that all are treated with equity and nobody is saved.

The equal protection paradox comes from P 11. Murphy describes it in this way:³

> If God (or any other rational being) shows mercy, then the mercy must not be arbitrary or capricious but must rather rest upon some good reason—some morally relevant feature of the situation that made the mercy seem appropriate…. Where reasons are given, it is possible to distinguish good from bad reasons, relevant from irrelevant ones. (“I showed him mercy because he was so sick” has a kind of sense lacking in “I showed him mercy because he was so handsome.”) But once a reason always a reason…. If I, as a rational being, showed mercy to Jones because of characteristic C, then it is presumably required of me (*rationally* required, not just morally required) that I show comparable mercy to C-bearing Smith.

Murphy is appealing to what we might call the *Principle of Equal Treatment*:

> If I treat person P in manner M for reason R, and if I have reason R with respect to P*, then I ought to treat P* in M as well.

This principle has the following corollary:

**Corollary:** For any two people P and P* who are similar because of some characteristic C, if C provides me with R for treating P in M, then it provides me with R for treating P* in M. So I ought to treat either both P and P* in M for R, or neither P nor P* in M for R.

---

So if I treat you mercifully because you have C, then I should show mercy to all with C. I should pardon all women accused of capital crimes, if I pardon one woman of a capital crime because she is a woman. Murphy thinks we ought to accept this principle, but he worries that this too will make mercy something obligatory, at least conditionally obligatory, along the lines of “I am never required to show mercy; but if I slip and show it even once, then I am rationally required to show it to all relevantly similar persons.”

Those who share Rawls’s intuition that justice simply is fairness should especially feel the bite of this problem. Were the state (or more broadly, the basic structure of society) to adopt Murphy’s conditional obligation view of mercy, it would have to outlaw mercy. Here’s why: Let’s assume that each act of mercy sets a precedent that in effect creates an exception or loophole in the law (e.g., everyone with condition C will get the merciful punishment if one does). Then (if we keep the distinction between equity and mercy) each act of mercy will serve to amend the law in the direction of being more lenient. Presumably, if the state wanted a more lenient law or a law with all sorts of loopholes, it could pass such a law. If the state instead has good reasons for passing the law without the loopholes or exceptions, then it could not permit its officials to act mercifully—under these circumstances, mercy is a force for lawlessness.

Those who try to analyze issues of justice in terms of competing rights also have a hard time with the equal protection paradox. Imagine person A (a judge or governor, e.g.)

---

4 Ibid.

5 The frequent public skirmishes between law-and-order attorneys general and lawmakers—who want to reduce the discretion of judges—and judges—who argue that judicial discretion is necessary to reach merciful and humane decisions—reflect the dynamic described here.
is given the authority and discretion to pardon one condemned prisoner a month. B and C are two condemned prisoners before A. We don’t want to say that B or C has a right to be treated mercifully (be pardoned, in this case), because that would imply an obligation on the part of A to pardon one or the other. If we describe this situation in terms of rights at all, we might say that A has the right to pardon or not. But we probably want to say that it is possible for A to exercise this right unjustly, as in the following examples:

1. A knows that B is guilty, but also that he is popular and his arrest and sentence were extremely unpopular with the public. A has doubts that C is really guilty, but C is unpopular with the crowds, who want him dead. A pardons B for the sake of social order, while C is executed.
2. B is beautiful, C is not, so A pardons B.
3. B promises A money and/or sexual favors if pardoned, C makes no such offers, so A pardons B.
4. A and B are of the same race, while C is of a different race, so A pardons B.
5. B is convicted of treason, while C is convicted of serial rape and murder. A flips a coin, and pardons C, who is released.

We don’t think very highly of A’s choices here. Yet if we say that A has a right to pardon one prisoner, so that he might not have pardoned anyone and instead pardons one person in each case, and if we say that the prisoners have no right to be pardoned, then it is hard to determine what the problem is. One might plausibly argue that by showing mercy at all, A is being merciful, and is therefore commendable insofar as being merciful is commendable. But we have serious problems with A’s reasons for exercising the right to be merciful. We want to say that the reasons are unjust, even if nobody’s rights are violated. It would take us too far afield to examine the dozens of rights-theories in detail, but these considerations do suggest that there is more to the idea of justice in this situation than can be captured by an idea of individual rights.

2. Anselm’s Solution to the Paradoxes
As we saw in chapter six, Anselm had great difficulty with these paradoxes in the *Proslogion*, the effects of which were made even greater by his theory of the divine attributes. While philosophers might find some way to wiggle out of these problems in the human situation, Anselm’s view that mercy and justice are both attributes of God requires him to define mercy and justice so that they are consistent with each other (solving the first paradox) and so that the exercise of mercy is governed by just principles (solving the second paradox). As we also saw, Anselm’s solved these problems by defining mercy and justice in terms of global rectitude or the right ordering of all creation. Anselm defines justice as rectitude of will for the sake of that rectitude itself, and, as we argued, to show mercy on his view is to attempt to restore the will of another to justice. Mercy has justice as its end, and so it is clearly consistent with justice—solving the first paradox. Anselm’s commitment to order gives him a standard by which to judge the exercise of mercy—mercy can be offered to all those who lack justice, until the point when doing so threatens due order. Anselm thus finds a principled reason to exercise mercy, which avoids the twin dangers of arbitrary judgments and of spiraling into lawlessness and excessive leniency.

Anselm’s solution is somewhat radical, though, because it dispenses with several commonly held views about punishment and justice. There is no place for retribution in this scheme (where retribution is defined as a duty to punish a wrongdoer just because she is a wrongdoer; that one is guilty is a necessary and sufficient condition for deserving punishment). On an Anselmian view, punishment and mercy are both justified in terms of whether they promote and defend due order. They do not make reference to the past actions of the wrongdoer, except insofar as those past actions affect due order. They do
make reference to the character of the wrongdoer, but the decision to punish is based both on whether or not the person is guilty of some crime, and also on whether the person is able to be restored to justice. On the Anselmian view, guilt is a necessary but not sufficient condition for punishment. To punish a person, it is necessary to determine not only that she is guilty, but also that not punishing her would endanger due order in some way.

9.3 Anselmian and Other Alternative Approaches to Punishment

As we saw above, several of Anselm’s political principles are concerned with punishment. Recall that Anselm subscribed to the following propositions:

- Punishing the presumptuous can help restore order.
- The duty to preserve order might compel a leader to punish.
- There is a point at which too much punishment decreases the order for which a leader is responsible.
- If punishment is necessary, the leader should punish reluctantly and not out of a joy of punishing.
- Ideally, the punishment will lead the punished to have a conversion, so that he develops a principled affection for the foundation of the regime.
- Even in contexts where punishment isn’t relevant, a leader still needs to bring others to a principled support of right order in the relevant sphere.

These last two principles follow from another principle, namely that one cannot rule stably simply by appealing to the self-interest of the ruled. Anselm rejects theories that justify punishment based solely on its manipulative or deterrence effect, arguing that rule by using such punishments could not be just.

Anselm holds that punishment is a central aspect of ruling, and so a just ruler must punish justly. When one’s subjects are affected by original sin and their personal sins and defects, ruling well requires ruling with mercy, since mercy helps restore the subject to justice. Those who have fallen short of justice do represent some threat to the
ruler, but if the ruler is strong enough, he can both subdue the wrongdoers and win them back to his loyalty. For Anselm, doing this where possible is an obligation of good government.

Anselm’s view of punishment is drawn from his experiences as a monk and abbot living under the Benedictine Rule. Chapter 27 insists that

> The abbot must take care, with diligence and cautious wisdom, not to lose any of his flock. He must remember that he has undertaken the care of sick souls, not tyranny over healthy ones. Let him fear the menacing words of the prophet, through whom spoke God, “What you saw to be fat you took to yourselves, and what was diseased you threw away” (Ezek. 34:3). He should follow the lead of the Good Shepherd who left ninety-nine sheep behind to search for the lost one. His compassion for weakness was such that he stooped to place the sheep on his shoulders to carry it back to the flock.

The abbot is not then to act as some sort monastic CEO, firing recalcitrant monks for underperformance. The disobedient monks are not to be treated as bad apples or sources of impurity in an otherwise healthy and holy group of men. All the monks are sinners and in need of healing; those who are disobedient are simply those who need the most urgent attention. So while the treatment of disobedient monks might seem harsh—permissible punishments range in severity from private reprimand, to public reprimand, to excommunication from the liturgy and from the common table, to enforced fasting or

---

6 Magnopere enim debet sollicitudinem gerere abbas et omni sagacitate et industria currere, ne aliquam de ovibus sibi creditis perdat. Noverit enim se infirmarum curam suscepisse animarum, non super sanas tyrannidem; et metuat prophetae comminationem per quam dicit Deus: *Quod crassum videbatis assumebatis et quod debile erat proiciebatis*. Et pastoris boni pium imitetur exemplum, qui, relictis nonaginta novem ovibus in montibus, abiti unam ovem quae erraverat quaerere; cuius infirmitati in tantum compassus est, ut eam in sacris humeris suis dignaretur imponere et sic reportare ad gregem.
flogging—these are all intended as medicine, as rehabilitative punishments, to win the monk (always called frater or brother, even if he has left the monastery) back to obedience and good will.

Contrast this with Rawls’s attitude towards the “unreasonable” living under the regime of justice as fairness. Rawls divides the liberal state into the reasonable and the unreasonable, and sees the unreasonable as a threat to the ideal state of strict compliance. As we have already examined at length, Rawls holds that, for any unreasonable people living under the ideal conditions of strict compliance in which justice as fairness makes sense, liberals have “the practical task of containing them—like war or disease—so that they do not overturn political justice.” The idea that the unreasonable are like a disease suggests that the reasonable are pure and healthy, and would be able to live according to justice were it not for the influence of the diseased members. This metaphor suggests why for Rawls’s view of strict compliance, then, it would be irresponsible to allow the unreasonable to grow in numbers or political strength, just as it would be irresponsible to allow a disease to corrupt a healthy body. If retributivism is characterized by the duty to punish the guilty, then we can understand why Rawls’s idea of strict compliance would lend itself to a retributivist view of punishment, as we saw in chapter four that it does.

Benedict regards all those under the Rule as unable to live a life of strict justice, and so insists that the monastery be governed through mercy and compassion, within the

---

7 PL 64n19.

limits set by the abbot’s concern for the common good. Punishment, indeed all the activities of the monastery, can thus be regarded as medicinal, trying to cure the monks of the tendency to injustice that follows upon original sin. Were someone who believed in the pervasiveness of sin or injustice to hold a retributivist view of punishment, in which it is a moral duty to punish the sinful and unjust, then it’s hard to see how one could avoid punishing everyone. Faced with such a proposition, one might sympathize with Boso’s terrified exclamation: “there is no way by which wretched, insignificant man may escape.”

Anselm’s view of punishment is neither retributivist nor based on deterrence, the two theories that dominated the philosophical discussion for most of the twentieth century. But it does have significant overlap with three alternative schools of punishment: the rehabilitative, moral education, and restorative justice approaches. Benedict’s imagery of punishment as medicine suggests that the Anselmian view might instead be a form of rehabilitation theory, the third major type of theory to win advocates in the last century. Its sense that the goal of punishment is moral reform has similarities with the moral education theories of punishment, which we briefly mentioned in chapter four. On the other hand, its concern for public order combined with its rejection of retributivism makes it sound like the vision of punishment proposed by advocates of restorative justice. In what follows, I hope to show that the Anselmian view does have affinities to these three views, but that it is an improvement over the most sophisticated versions of these

9 CDH I. 24: “non est qua evadat miser homuncio”
views developed so far. As we will see, it is closest to the restorative justice view, though different in some significant ways.

9.4 Is Anselmian punishment a rehabilitative theory?

In chapter four, we briefly mentioned the rehabilitation theory of punishment, which was so popular in the middle decades of last century. Barbara Wootton was perhaps the most famous advocate of this theory, which is based on the premise that criminality is often a sign of mental illness. Sociological studies in the years immediately after World War II seemed to suggest that class and economic factors did not play the role in criminal behavior that many expected, while the categories of mental illness and health promised to provide more adequate explanations. As Wootton writes, “We prefer today to analyze the infected individual rather than to eliminate the infection from the environment.” She defends this preference on several grounds—it seems easier to rehabilitate one person than to change all of society; the rehabilitation theory seems more humane because the treatment is tailored to the needs of the individual; rehabilitation of the few is likely to prove more cost effective and have fewer side effects than wide-scale social engineering.

Such rehabilitation theories were usually rejected, for a variety of reasons. From the beginning, retributivists did not like the theories, because they rejected the idea of

10 Cf. Barbara Wootton, Social Science and Social Pathology (George Allen & Unwin, 1959).

11 Ibid., p. 329.

12 Cf. ibid., Chapter XI.
Indeed, not only did they reject the idea of desert, but they also recommended getting rid of standardized sentences for each crime, so that all people who committed the same crime would get roughly the same sentence. Wootton thought it reasonable that, once guilt has been determined, all sentences should be indeterminate.\textsuperscript{13}

Modern reformative methods of treatment, whether based upon psychiatric diagnoses or upon statistical prediction tables, demand that similar offences should entail very dissimilar consequences. Where two men are found guilty of the same offence, perhaps even on a joint charge, differences in personality may indicate that one is likely to respond favorably to lenient treatment and sympathetic understanding, whilst the other is a bad risk for whom prolonged detention is the only safe course….To a public conditioned to the belief that the punishment should fit the crime, such inequalities of treatment can only appear as monstrous injustices.

As Wootton predicted, political opponents of this sort of theory were able to defeat this suggestion by appealing straightforwardly to the widespread belief that people should go to prison only for a set time.\textsuperscript{14} From a philosophical perspective, however the political defeat isn’t decisive, since it is based on a question-begging assumption, namely, that sentences should always be determinate.\textsuperscript{15}

\textsuperscript{13} Ibid., pp. 335-6.


\textsuperscript{15} The political defeat was not politically decisive either, since at present some states have adopted indefinite incarceration as a policy to deal with certain sorts of criminals—sex offenders, typically—who seem likely to commit violent crimes were they to be released. Cf. Laura Mansnerus, “Sex-Offender Release Policy Faces Change in New Jersey,” \textit{The New York Times}, January 26, 2004, p. B5, regarding appellate challenges to New Jersey’s “Sexually Violent Predator’s Act,” one of sixteen state laws commonly known as “Megan’s Laws.” In \textit{Kansas v. Hendricks} (1997), the Supreme Court of the United States upheld such laws as constitutional.
Those interested in crime control were at first attracted by the promise of the rehabilitation approaches to be able to reduce recidivism. Their enthusiasm waned, however, once the rehabilitation policies were enacted and decades-long crime waves followed. “In the 1960s and early 1970s, we witnessed the rise and fall of the rehabilitative ideal. Studies showed, or were interpreted to show, that participation in treatment programs ranging from psychological counseling to remedial education had little or not effect on convicts’ postinstitutional behavior.”16 Today, the view that criminality as such is a curable mental illness is fairly discredited.17 Since the rehabilitationist model had justified the practice of giving judges wide latitude in sentencing, the rejection of that model has justified the current trend towards stricter sentencing guidelines.

Andrew Von Hirsch, in his helpful summary of the debates over sentencing,18 notes that the decline of the rehabilitation model was due in part to a realization that there are limits to the malleability of prisoners.19 The science of psychiatry declined in prestige among policy makers as the predictions of Wooton and others about the power of social science proved overly optimistic. Von Hirsch also points out that the rehabilitationist

16 John J. DiIulio, Jr., No Escape: The Future of American Corrections. (Basic Books, 1991) p. 7. See also pp. 126-147 for an assessment of Patuxent Institution, a.k.a. the “Clockwork Orange” prison, Maryland’s experimental prison based on just such theories.

17 Cf. ibid., p. 104.


19 Ibid., pp.4-5.
vision was tied up with another idea that has also come into disrepute: the idea of *predictive restraint*, that “sentencing and correctional officials were supposed to gauge not only individual offenders’ treatment needs, but also their likelihood of returning to crime.”

Von Hirsch notes that the rehabilitation approaches to punishment nearly always made reference to this other element. He argues persuasively that it was the two elements together that made the rehabilitationist approach so appealing—“therapy could always be tried on apparently amenable defendants, but always with a fail-safe: the offender who seemed unsuitable for or unresponsive to treatment could be separated from the community,” that is, could be retained in prison indefinitely. This view has few adherents any more, for two reasons: consequentialists abandoned it when studies suggested that the social scientists and psychiatrists were not terribly good at predicting which inmates would recidivate, while others, on liberalism of fear grounds, increasingly came to worry that government and medical bureaucrats were likely to commit great cruelties in the name of humanity.

Today, a much-chastened version of rehabilitationist punishment is making a comeback. Having discarded the extravagant claims of mid-century psychiatry, rehabilitation theories are gaining adherents, in large part by showing that at least some well-designed programs do work to reduce recidivism, as evidenced by rigorous studies. John DiIulio suggests that there three different purposes have been advanced for

---

20 Ibid., p. 5.

21 Ibid.

having correctional programs: 1) they contribute to the rehabilitation of the offender; 2) they facilitate the operation of safe prisons by giving the inmates and staff something to do; 3) they are an essential component of humane custody. The second purpose is the one most persuasive to corrections officers, and therefore the most likely to survive despite policy shifts. The third view is perhaps the most controversial, with some people more or less equating corrections programs with humane punishment, while others view it as a moral outrage that convicted felons can get educational, employment, and other opportunities not always available to law-abiding citizens.

The view most persuasive to legislators, though, seems to be the first view—if corrections programs can reduce recidivism, they are worth the cost and the hassle; otherwise, it is hard to justify them to their constituents. Given our larger concerns, it would seem that this is the most sensible view. Programs that actually reduce recidivism promote stability, since they take someone who chooses to break the law and convince him to choose to follow the law. They show mercy to the criminal, which fosters support for the regime among the entire population by expressing the state’s commitment to its citizens. And if they are carefully controlled (no unsupervised weekend furloughs for violent offenders, e.g.), they can be consistent with the incapacitation and social defense aspects of punishment. While the Anselmian view of the criminal is similar to that of the

23 DiIulio (No Escape, p. 116) quotes one officer who defends corrections programs with rather salty language: “I don’t see how the hell you could run a joint without programs. Without them, staff would get fed up, and the convicts would probably be more trouble to themselves and everybody else...There was no way we were going to f—k up everything just because some asshole legislator or double asshole professor decided that rehabilitation was dead. Rehabilitation isn’t in the vocabulary, never was. [Programs] are like custody, like food service...They are part of what it takes to run a good joint, period.”
discredited psychiatric version of rehabilitation, since both views hold that prisoners are in some way defective or sick, the Anselmian response to this illness is quite different. Since on the Anselmian view the illness is explicitly moral rather than psychological, the illness is not cured until the will is healthy; correct behavior is not enough. The goal is for the criminals to repent and decide to obey the law voluntarily. And since the Anselmian view acknowledges that persons cannot be compelled to will something unwillingly, it is prepared to accept that despite our best efforts some people will choose not to comply with the law, that people are not always malleable. An Anselmian, thus, will not argue for indefinite incarceration on the grounds that if a criminal isn’t cured, more treatment will help. It does leave the door open for indefinite incarceration on other grounds, though, which we will discuss more below.

In sum, the Anselmian theory is interested in the rehabilitation of the prisoner, but it approaches rehabilitation from a different view of the human person than the heavily criticized rehabilitationist approaches of the mid twentieth century. Rather than relying on the behavior modification techniques of psychiatry or the different but no less manipulative deterrent approaches to punishment, the Anselmian theory attempts a moral rehabilitation that acknowledges the absolute freedom of the human will. This view sets the bar much higher than the other views, because on this account obeying the law out of self-interested fear does not count as rehabilitation. On the Anselmian view a criminal is rehabilitated only if he has a principled affection for the values of the society, and agrees not to break the law for the reason that it is wrong to break the law. This understanding of ‘rehabilitation’ resembles quite closely the moral education approach to punishment we examined earlier.
9.5 Is Anselmian punishment a moral education theory?

In chapter four we sketched the main argument of the moral education approach to punishment. These theories have a descriptive and normative component: they set out to explain why many aspects of our practice of punishment make sense, and then they offer proposals for reforming the practice to make it better fit the theory. On this view, punishment communicates a complex message about power relations, community values, and the appropriate place of a citizen in society. At least, we suggested in chapter four, punishment sends the following message:

the power in authority P believes strongly that the act A by the subject of punishment S is wrong; that S is responsible for choosing to A, knowing that it is wrong; that the rationale V underlying the law L which S violated is an objectively valuable good; and that S’s decision to A makes manifest that S does not appreciate the good of V.

If P is able to make the punishment fit the crime, the punishment also communicates to S that “you did this to your victim; you hate it, so imagine how your victim feels.” The hope is that S will realize just how seriously others take V, and perhaps will learn to adopt the point of view of their victim, an important stage of repentance.

If we take what Anselm says about justice and mercy for God and generalize it to apply to any political leader, then we can see that Anselm has a moral education theory of punishment, indeed, a moral education theory of state power more generally. Anselm believes that punishment must preserve and protect the general political order, but that it does this best when it takes those who try to usurp that order and instills in them a respect for the principles upon which the order is based. Anselm assumes the criminals are responsible for their actions, and that either they reject the justice of the political order or
they do not care about justice. It is therefore important for the stability and general
flourishing of society that an attempt be made to restore in the criminals a will to preserve
justice for its own sake. Punishment should not undermine the allegiance of the general
public towards the regime, either by being ineffective or by being unjust, and so it has to
take into account not just the education of the criminal but also the education and safety
of the public at large. But stability and good order demand that where it is possible the
criminals be offered mercy within the constraints of justice, so that they might be
rehabilitated. Thus, Anselm’s theory is an example of a moral education theory of
punishment.

In a discussion of Herbert Morris’s version of a moral education theory, R. A.
Duff and David Garland advance a particularly clear objection to such theories. They
write that the problem is that this theory seems anti-liberal, communitarian perhaps. 24

It is not at all clear that our societies constitute the kinds of
genuine moral community, united by shared values and mutual
concern, within which both crime and punishment could have the
moral character that Morris ascribes to them. … But we must also
ask whether Morris offers a plausible ideal of punishment. In
particular, should it be the state’s job to take such an interest in the
moral or spiritual well-being of its citizens? Would we really want
the state, even in a better society than our own, to take on the role
of parent, or of abbot, seeking to bring (by such coercive measures
as punishment) its wayward members to repentance and reform?
Such a view of the state’s proper role is radically at odds with any
liberal conception of the state, and presupposes some more
communitarian political theory.

Punishment’” in A Reader on Punishment, R.A. Duff and David Garland, ed. (Oxford
If the goal of a moral education theory is to inculcate a respect for the principles of the regime, it assumes that there are some principles that are widely accepted. But what happens if, as certainly might happen in a pluralistic society, the criminal doesn’t share those principles for perfectly rational reasons? If our goal is to convert them, do we punish until they convert, perhaps, indefinitely? If they were ‘hardened criminals,’ is it wrong to use particularly harsh punishments to soften them up? What happens if some prisoner does repent and convert before her sentence is up—do we release her early? Finally, isn’t there something inherently egalitarian about liberal democracies that is violated when the state is considered to be a moral educator on the model of a parent?

Before we respond to these objections, we should note that they move quickly from discussing the theory of punishment in itself to asking questions about how it fits into a larger theory of the state. This reinforces the point we made in chapter four that punishment is a central activity of any state, and so it is very important that a theory of the state have a good theory of punishment. Moral education theorists argue that not only ought the state to have a good theory of punishment, it also needs some account of what counts as moral, since punishment (as distinct from mere regulatory fines, such as parking tickets, which don’t seem to be punishments properly speaking) cannot help but teach some sort of morality. In this they side with Joel Feinberg, who noted that as a matter of pure description, the decision to punish a certain behavior communicates several messages, all of which affirm that the behavior is in some way bad or wrong.\(^{25}\)

For sure, the moral education theory is not simply descriptive but is also normative,

---

arguing that because punishment has an educative function anyway, one ought to design the penal system to take advantage of this fact. But implicit in the normative claim is the descriptive account of punishment as educative.

This is not to say that before adopting a theory of punishment one must settle every moral issue. As Hampton points out, there is a difference between the question of what one punishes and how one punishes, and the moral education theory at least only addresses the second question.\(^2\) It is compatible with different theories about what areas should be punished: one can agree or disagree with Mill about whether ‘victimless crimes’ should be penalized, without affecting the claim that the law and penalties communicate what society thinks is right. Hampton suggests that some people might grant this but argue that the state should only presume to educate serious felons. To this, she has a ready response, “Is this not exactly the kind of sentiment behind the libertarians’ call for extensive constraints on the state’s role and power?”\(^2\) This suggests that, although the questions of what one punishes and how one punishes are distinct, they are not entirely independent. Political compromises in a divided and pluralistic society may lead to a very narrow view of what should be punished, as the libertarians hope. On the other hand, political consensus (however short term) might lead to the criminalization of behaviors (the use of marijuana, e.g.) that some people don’t want criminalized or to the decriminalization of behaviors (sodomy, e.g.) that some people want criminalized.


\(^2\) Ibid., p. 219.
These larger political consequences are not unique to the moral education theory of punishment—questions about what to punish are likely to be politically and morally controversial no matter what the theory of how to punish. But the moral education theory does put pressure on the state to abandon any pretenses to moral neutrality. If punishment always serves to connect the criminal with the moral values underlying the law, then the state cannot pretend that in its exercise of power it does not advance moral claims, at least some of the time. Nor can it pretend that those moral claims will be universally accepted, unless it is also prepared to hold that there will not be any criminals. So the state has to decide what moral principles it will promote even over the objections of some of its citizens.

If the moral education theorists are right about this descriptive claim, then they have a response to the objection noted above. Every regime that punishes (i.e., every regime) imposes some moral principles on its citizens. The state has the option of taking responsibility for this feature of punishing by taking an interest in the moral and spiritual good of its citizens, or of ducking this sort of responsibility, and by its disinterest authoritatively teaching the view that the moral and spiritual good of its citizens is not valuable. It’s not outrageous to think that liberalism ought to embrace the first position.

Moral education theorists note that their theory makes a positive moral claim against deterrence theories, namely, that the good of the wrongdoer is a primary concern of punishment. It is not the only concern, for they also believe that the promulgation and enforcement of the law educates the public about how important these values are to society, if less directly than punishment educates the criminal. But their additional concern about the good of the criminal puts certain requirements on the practice of
punishment. Execution of the offender is obviously incompatible with his moral education, for example, no matter the deterrent effect. The same is true for torture or other cruel punishments. A deterrence theorist might argue that prison serves as a deterrent only to someone for whom it is worse to be imprisoned than not. One cannot be threatened by prison if going to prison is an improvement in one’s living conditions. Thus, many deterrence theorists conclude that prisons should be at least slightly less pleasant than the environment from which the worst criminals tend to come. On the other hand, a moral education theorist would instead have the goal that the prisoner be punished, but not treated vindictively, or in ways not in accord with their human dignity. J. D. Mabbot’s classic article “Punishment” suggests that criminals ought to be released from jail in “a reasonable frame of mind” thinking something like: “‘Well, I’ve done my time. They were not too bad to me. Prison is prison and not a bed of roses. Still, they didn’t rub it in…’ This ‘reasonable state of mind’ is one in which a prisoner on release feels he has been punished but not additionally insulted or ill-treated.”

As Morris argues, it is mistaken to think that paternalism is the enemy of autonomy. Parents are not indifferent to the moral autonomy of their children. Rather, they punish because they are concerned that their children come to internalize both what is the correct behavior and why it is the correct behavior. They assume their children are capable of free choice and of moral responsibility, and that they can learn from their moral mistakes. Similarly, laws and sanctions communicate what society holds to be permissible and impermissible behaviors, how seriously it regards different goods and

28 J. D. Mabbot, “Punishment” in Acton, p. 53.
different wrongs, and what is the nature of the evil proscribed (especially if it is possible for the punishment to ‘fit the crime’ in its severity and character). In teaching this, they don’t demean the criminals, but rather treat them as morally responsible and as capable of rational reflection. It is possible that a person commits a crime because she disagrees with what society thinks is wrong or with how wrong society thinks it is, and is willing to risk being punished in order to ask society to reexamine its values and practices. This is how we characterize civil disobedience. But most crimes ought not be characterized as civil disobedience, surely; and even where a crime is done for reasons of civil disobedience, it does not for that reason mean that society has to agree to change. Hampton notes that we don’t respect the rapist’s protest, “Who is the state to tell me that my rape of this woman is immoral?”, because his actions “betray a serious inability to make decisions about immoral and moral actions.” Nor is it the case that every act of civil disobedience should be taken as seriously as that of Martin Luther King, Jr. (We will say more about civil disobedience in a little bit.)

In his intelligent and often persuasive book Trials and Punishments, R. A. Duff embeds the moral education justification of punishment within a moral education analysis of trials and sentencing. Duff is committed to the view that “the law should respect every citizen as a rational and autonomous agent,” and sees moral education theories as the only justification of punishment consistent with this conviction. The criminal law has as


among its aims and principles the communication to the offender that his behavior is blameworthy and merits moral censure. We conduct trials in the way that we do in part as a way to fix blame on the defender and to make him consider the community’s judgment that he has behaved wrongly. Sentencing, too, is among other things an act of blaming, and thus is an attempt at moral persuasion and education.\footnote{Duff, \textit{Trials and Punishments}, esp. ch. 4.} Given this analysis, Duff asks, why must we punish at all, if we have already blamed the criminal in the trial and sentencing phases? If the other two parts of the criminal justice system are successful, how does the hard treatment of punishment add anything to the education of the prisoner? This question is very important on Duff’s analysis, since given his view that the law must respect the autonomy of those it punishes, anything that smacks of coercion needs justification.

Duff justifies the punishment of criminals in a secular state by analogy with how a church disciplines the sinners within its community. Sin separates the sinner from God, and also from the rest of the church.\footnote{Ibid., p. 248.} Since the church is a community of believers, bound together by their love of God and by their spiritual relationship to each other in God, my impenitent sinfulness must also injure my relationship with the church and with my fellows in the church: it separates me not just from God but from them.

Since the church is wounded by the sin of its members, it may thus take some sort of action to communicate this wound to the sinner, “by excluding [him] from its services, its

\begin{flushright}
\footnote{Duff, \textit{Trials and Punishments}, esp. ch. 4.}  
\footnote{Ibid., p. 248.}
\end{flushright}
holy places, or even from contact and communion with its members.” Yet the church never simply rejects the excommunicant. Rather, “he must be allowed, encouraged and helped to restore himself through penance, to a full communion with God and church; he was still to be treated and cared for as a member of the community.”

Sometimes, the church seeks to impose a penance on the sinner, which she might not accept. In what appears to be a thought experiment (or perhaps a vague gesture at ecclesiastical practices of yore), Duff supposes that the church decides to force the sinner to undertake the penance, “that she is then forcibly subjected to solitary confinement, to a diet of bread and water, or even to physical beatings.” Duff thinks that such forcible penance could be justified, but only after four conditions are satisfied:

i. The sinner must be seen as a member of the community
ii. The sin must be seen to separate her from the community and her own good
iii. In imposing the penance, the aim must be to bring about repentance and reconciliation, and thus her own good
iv. The aim of the coercion is help her understand the coercion as a penance and to assent to it as such.

Duff argues that only if the sinner comes to accept and understand the coercion as a penance is it actually a penance. “If the suffering to which [the sinner] is subjected is imposed on him in the right way and in the right spirit it can bring home to him, as merely verbal or symbolic condemnation would not, the extent and nature of his sin.”

---

34 Ibid.
35 Ibid., p. 249.
36 Ibid., p. 250.
37 Ibid., pp. 250-1.
38 Ibid., p. 251.
But if done in this way, the coercion addresses the sinner as a rational moral agent, and
does not try to “coerce or manipulate his will”—say through torture or extreme
psychological pressure.

Duff thinks that one can defend and justify the punishment of criminals as an
analogous sort of compulsory penance.\textsuperscript{39} He proposes that we conceive of the state as a
moral community whose members share a regard for certain central values and for the
other members of the community. If the laws of the community are just, then a crime
against the laws separates the criminal both from the community and from the values that
make the community able to contribute to his well being. Repentant criminals see the
damage that their crimes have done to the community and to themselves, but unrepentant
criminals need help to see this. If punishment is to provide this help, it must communicate
to the criminal that his crime was wrong, it must bring him to repent of that crime, and
thus to will the punishment as a penance which can atone for the crime. There will be
some criminals who do not repent, and thus the punishment does not achieve its aim. Yet,
this sort of fallibility is essential to the justification of punishment, since the possibility of
failure shows that the law respects the criminal as a free and rational being who ought not
be coerced into repenting.

Duff considers it a strength of his theory that it sees the value of punishment in
the punishment itself, and not in the opportunity it gives for psychiatrists to provide
reformative treatment nor in the efficiency of its technique for changing the behavior and
motivation of criminals. It is the punishment itself which, by becoming a penance, is to

\textsuperscript{39} Ibid., p. 254ff.
cure the soul of the criminal. Duff even goes so far as to say that “a sane offender has a right to be punished...rather than be ignored or dismissed...We owe it to him, as well as to those he has injured, to condemn his crime and to try to bring him to repentance or reform.”

Duff summarizes the view:

Punishment on is, on this conception, communicative, retributive, and reformative. It is communicative in that it seeks to communicate to the criminal a proper understanding of his crime: by imposing on him some material injury which can be seen as injurious even through the eyes of egoistical self-interest, we hope to represent, and to force on his attention, the harm he has done both to others and to himself; by imprisonment...we give material and symbolic expression to the spiritual separation created by his crime. It is retributive in that it is a response to his past crime whose meaning lies in the judgment on, and the condemnation of, that crime which it seeks to communicate; and whose aim is to bring the criminal to suffer the pain of remorse which he ought to feel when he considers his crime. And it is reformative in that it aims to bring the criminal through his understanding of his crime and his punishment, to repent his crime; to accept his punishment as an appropriate vehicle for his repentance; and thus to reform himself—to reconcile himself to the values which his crime denied and to the community to which he belongs.

An Anselmian will obviously find much that is attractive in this view. Indeed, given its appeal to penitential atonement as practiced in Christian churches, and the influence of Anselm’s theory of the atonement within Christianity, it is not unlikely that there is a genealogical connection between Duff’s view and Anselm’s. But Duff’s view is susceptible to some criticisms that Anselm’s is not. Duff’s argument seems to prove too much, making it hard to justify using alternatives to punishment. If punishment is always

---

40 Ibid., p. 264.

good for the criminal, if the criminal has a right to be educated through punishment, then it would be wrong to deny the criminal such punishment. As we saw above, Duff thinks that the pain of punishment itself can help the wrongdoer atone for his crimes, and that by failing to punish we fail to help the prisoner come closer to his own good. But advocates of alternatives to punishment argue that there could be other ways a repentant prisoner can repair the damage done by the crime—e.g., community service, making restitution, ‘conferencing’ with the victims of his crime in an attempt to rebuild the relationships his crime had ruptured. If parole and electronic surveillance methods help the prisoner reintegrate into society, should they be rejected because they are insufficiently penitential? If Duff’s argument prohibits him from adopting alternatives to punishment, then the retributive elements of Duff’s view loom larger. But as we’ve argued, if justice is thought to entail that punishment of criminals is a duty, then justice is defined so that mercy is unjust, with all the destabilizing and inhumane consequences we have discussed at length. Anselm’s view, on the other hand, is open to alternatives to punishment, so long as they are likely to foster a just will among the wrongdoers while not posing too much of a danger to the social order.

A second criticism of Duff’s view holds that it is simply impossible to combine a medicinal view of punishment as the healing of a sick soul with the view that every citizen ought to be governed only by those laws he finds rationally compelling. The Kantian account of autonomy that underwrites the second view was designed to cut off just the sort of paternalistic approach that the first view suggests. Using heteronomy to foster autonomy in an adult seems to be about as paradoxical as using squareness to foster roundness—threatening people with coercive punishment unless they develop a
principled affection for the ideal of autonomy seems likely to undermine the very autonomy the regime ultimately wants to promote. Anselm’s view, which does not make reference to autonomy, is not vulnerable to this criticism.

A related criticism notes that Duff’s overall account in *Trials and Punishments* assumes a communitarian view of the state, in which all members of the community share at least some of the same values, and it is these values that justify the quality of moral censure inherent in trials, sentencing, and punishment. This view is not unlike Rawls’s ‘overlapping consensus’ view, and is susceptible to some of the same criticisms we advanced against Rawls’s view of punishment: it might work in ideal conditions, but not when people do not subscribe to the values expressed in the law for good and defensible reasons. As Wesley Cragg points out, if the state cannot defend the superior rationality of the value behind the punishment, then it cannot be engaged in rational persuasion, and therefore it continues to punish either out of a sense that punishment is good in itself (i.e., from retribution) or out of the hope that the criminal might decide to obey the law for prudential, non-rational means, in which case it would simply be deterrent punishment.

---

42 This assumes that every viable regime will impose punishments for mala prohibita as well as for mala in se. A crime is a malum in se if there is something about the nature of the crime that makes it bad in itself, as in larceny, rape, and murder. A crime is a malum prohibitum if it is illegal only because it violates some statute. For any crime that is a malum in se, there will not be a rational argument in defense of it (by definition). However, this need not be the case for mala prohibita (again, by definition).

Anselm’s view, on the other hand, justifies punishment as the attempt to subdue the presumptuous for the sake of due order. The communicative and persuasive functions of trials, sentencing, and punishment are directed towards the right ordering of society, and are key aspects of promoting the stability of the regime. If the criminal can be won over, the regime will be more stable. But Anselmian punishment need not try to persuade the offender that she has acted *irrationally* in rejecting some contentious community value. For it to be justified, it need only communicate that the offender acted *presumptuously* by breaking a law it was not her place to break. One does not need some complicated theory to explain how punishment might do this—the very act of legal punishment communicates at least that the state regards the criminal’s act as against the law, and the law sets limits to what behavior is permitted and what is presumptuous. Nor does the Anselmian view require that in a pluralistic society there be one moral view regarding crimes that are *malum prohibitum*. Under the rules of democracies, it is always presumptuous to step outside the law to assert one’s own will. All the Anselmian requires is that citizens submit to the rules (and roles) as given. If you want to change something, run for office, don’t break the law.

A fourth criticism notes that Duff’s view seems to overemphasize the role of trials and sentencing as attempts at moral suasion. There is much that is insightful in Duff’s account of the trial process, but it seems that persuading the criminal that the law is morally just is at best a secondary aim, which the judge is at liberty not to pursue. On the other hand, the primary aim of the trial is to establish a question of fact, and that the primary aim of sentencing is to apply the law, and these aims must be carried out in each trial. The judge and jury are not free to reconsider the merits of the law or the moral
quality of the culprit’s actions—indeed, laws against jury nullification are designed to cut off appeals to extra-legal moral intuitions. Sometimes the law is clear where the moral issue is controversial. Duff considers this objection, and tries to blunt its force by arguing that under the doctrine of separation of powers, the court has to assume that the laws it applies are morally justified, because they have been passed by the legislature. Yet, this response misses the point. The legislature is not obligated to provide a moral justification for its laws, and the court is not obligated to believe that the laws have some moral justification, let alone to believe that the procedures by which a bill becomes law somehow bestow moral justification upon the substance of the law. The procedures by which a bill becomes a law do justify the law, but only formally, insofar as the law is a law; the justification need not extend to its content.

Again, the Anselmian account of crime as presumption provides the moral justification for obeying any law that has been legitimately passed. Any criminal, as such, has acted presumptuously, in a way that is not permitted to someone in her role. Law and morality necessarily meet at this point—the criminal, in breaking the law, has committed an act of presumption, which is morally blameworthy. A person might have a view of justice which conflicts with the substance of the law, but that alone is insufficient moral

44 Duff, *Trials and Punishments*, pp. 137-8. “A criminal trial in a system like our own must presuppose, not merely the fact of such a separation of powers, but the availability of an adequate moral justification for the secondary rules which thus limit the court’s responsibilities and require the citizen to obey the law until it is duly changed; the existence of a forum in which this justification can be offered and criticized; and the availability of procedures through which the citizen can try to change the law. Only then can the claim, on which the trial itself depends, be maintained that the defendant has an obligation to obey the law, and to accept the authority of the court.”

justification for breaking the law. One would have the necessary moral justification only if one can plausibly argue that to violate the law on this occasion contributes to the social order rather than detracts from it, that it is indeed one’s place to do so—that it is not presumptuous. In that case, the lawbreaker has declared war on the state, and the state has an obligation in self-defense to subdue the lawbreaker. Of course, any act of punishing gives the punisher a chance to evaluate whether the punishment advances the stability of the regime: Is the value behind the law great enough to justify the expense and effort of subduing and then rehabilitating this citizen? Is the law still a good law to have? Will this act of punishment undermine support for the regime more generally? It is this feature of punishment, viz., that it forces the regime to reconsider the reasons and values behind the law, which makes civil disobedience such a powerful tool for legal reform. In successful civil disobedience, the offender convinces the lawmakers (perhaps unintentionally) that the current law decreases the stability of the social order, and so they alter the law to make it more stable. An Anselmian, though, does not need a separate theory to handle civil disobedience, since on this view all lawbreaking (at least, of laws against mala prohibita) is civil disobedience.

In sum, while Anselm’s view is a moral education theory, by holding that presumption is morally blameworthy it is able to explain why the state can hold that all lawbreaking is immoral without having to assume that pluralistic societies agree on more substantive moral claims. Unlike most moral education theories, it is at home in morally pluralistic societies, so long as they are governed according to the rule of law. Yet it shares the great strength of all such theories in that it acknowledges and explains the moral dimension to legal punishment, and thus to the law itself.
9.6 Is Anselmian Justice a Restorative Justice Theory?

‘Restorative justice’ is a name given to a diverse set of initiatives where the victims, offenders, and communities affected by a particular crime collectively decide how to deal with the consequences of the offence, its implications for the state of future relationships in and with the community, and any other aspect of social life that matters to members of the community affected by the crime. Because restorative justice is the name of a movement rather than one philosophical position among others, it can be hard to define, but most practitioners seem to agree on a few elements: The state is not the primary agent of restorative justice, but the stakeholders are—the victims, offenders, and communities. The aim of restorative justice is the subjective satisfaction of the stakeholders in a particular case, and not some objective goal of justice or equal treatment or due process or retribution. Hence, restorative justice tries to restore relationships in the community, and foster the values of healing, respectful dialogue, responsibility, moral learning, and forgiveness. Advocates of this approach argue that restorative justice is a moral improvement over retributive approaches to criminal justice, both for the offender and for the victim. They also claim, with the support of some preliminary studies, that restorative justice techniques can reduce recidivism and save the costs associated with incarceration. Because of its emphasis on restoring order to relationships damaged by crime, it bears sufficient similarities to Anselmian justice that it is worth discussing

---


47 See John Braithwaite, Restorative Justice and Responsive Regulation (Oxford University Press, 2002).
whether Anselmian justice is a form of restorative justice, or at least whether it is compatible with elements of restorative justice.

If the claims about recidivism and crime prevention can be borne out over time, consequentialists have every reason to support the restorative justice movement. But other philosophers have raised problems with the principles of the restorative justice approach. Andrew Ashworth arranges many of these objections under four headings:  

I. *The state should retain responsibility for the administration of criminal justice.* Proponents of RJ show a naïve preference for informal procedures of criminal justice over formal ones. The public (in the form of the state) ought to have control over criminal justice to prevent vigilantism.

II. *An independent and impartial tribunal should decide the outcome of criminal cases.* By having the power of sentencing delegated to the local community, RJ leads to a ‘ZIP code lottery,’ where the standards of justice vary with geographical location. Insofar as RJ involves the victim in sentencing, it detracts from the impartiality and the independence of the criminal justice system. There’s an additional worry that the victims might want disproportionately harsh punishment. That the police are often involved in the RJ process (including the sentencing process) creates a separation of powers worry that the police are investigator, prosecutor, judge, and jury.

III. *The principles applied by such a tribunal ought to be consistent with the rule of law and proportionate to the seriousness of the offences.* RJ violates the principle that similar offenders and similar offences should be treated similarly. It works against the rule of law because it focuses on the offender and not on the demands of justice. It seems at least as difficult to specify how much ‘restoration’ will be accomplished by a particular sentence as it is to specify how much ‘retribution’ a sentence accomplishes.

IV. *Victims have a legitimate interest in reparation or compensation from the offender but not in the making of any wider orders or sentences.* Victims, as victims, have at most a right to compensation from them for the harm done; their status as victims does not grant them an additional role in the enforcement of justice. On the other hand, their interest in the punishment of the criminal is no greater than that of any other citizen, because punishment is a response to an offence against all the citizens.

---

48 Andrew Ashworth, “Is Restorative Justice the Way Forward for Criminal Justice?”
Several of these objections reflect a desert theory of punishment and a Rawlsian social contract picture of the state. Wesley Cragg, however, has developed the outlines of a theory of restorative justice that avoids most of these worries, and which bears important similarities to the Anselmian view of punishment we’ve been developing here. Examining this theory will help us bring out important features of the Anselmian account, especially how it can justify the use of violence in punishment given its emphasis on mercy.

Cragg’s theory begins with the proposal that the purpose of the law is to resolve conflicts in a way that is fair and which minimizes violence. Making conflict resolution the justifying aim of the law involves a significant break from the consequentialist view that the purpose of the law is to deter crime or from the view that the law is supposed to instantiate the principles of a just social contract. On Cragg’s view, legal systems are formal systems for resolving disputes, which can and do coexist with informal systems for resolving disputes. The formal systems are necessary because the informal systems, while often effective, can be unfair or ineffective in resolving disputes, and might even spin off into violence. Cragg argues that the decision whether to use formal or informal systems of conflict resolution ought to be determined by which system is most likely to reduce the use of force in each case. In a pluralistic society that lacks one set of community values, the formal systems will often be preferable, but if there is significant agreement about what can be done in a particular case, the informal conflict resolutions

---

would normally be better. Recourse to the law is justified when “on balance, given available alternatives, to have a legal system in place would reduce the vulnerability of those under its authority to the justified use of force and thereby provide improved opportunities for cooperation in the pursuit by individuals of their various goals.”

Given this approach, law enforcement officials ought to seek to minimize the occasions when they use coercion, engaging in coercion reluctantly, as a last resort, only when other methods of reducing conflict seem unpromising. Cragg suggests that the best law enforcement officers actually do this, exercising wide latitude in deciding whether an arrest, a show of force, or a mere caution will best achieve the goals of cooperation and societal protection. Indeed, only if we understand the goals of the police to be conflict resolution with a minimal use of force do we explain several perfectly ordinary activities of the police, such as traffic control, managing domestic disputes, and breaking up fights between juveniles without making arrests. If the purpose of the law is to prevent, manage, and resolve conflict so that it minimizes the use of force, then one would expect that law enforcement officials would prefer persuading rather than forcing people to obey the law—and often enough the actions of the police do just that.

This view of the law and its enforcement avoids several of the objections above. Cragg does not have the naïve belief that informal systems are better than formal; indeed he sees the avoidance of vigilantism as a key justification for the law at all. He thus

---


51 Cragg, *The Practice of Punishment*, ch. 5.

avoids the first objection. He can defend his view against the third objection by denying its assumptions about the role of legal institutions. If the rule of law is just one among many methods of conflict resolution, one that works better in some circumstances and worse in others, then there is no harm in using an informal procedure if it is found to be effective. If the conflict is resolved in a relatively stable fashion, then the method of conflict resolution is justified. The goal of conflict resolution helps to define what is meant by ‘restoration,’ so that the relationships damaged by an offence are ‘restored’ when all the parties lay aside their conflicts and their desires for revenge, and promise to obey the law.

Cragg thinks that sentencing can also be seen as a form of conflict resolution. Once a crime has been committed, conflict is the result, and the goals of the trial and sentencing process are to keep the conflict from escalating and to defuse it as much as possible. Criminal disputes are different from private law disputes because of their public character—the public has an important stake in the existence of the formal systems of conflict resolution, and direct flouting of the law serves to undermine the law if it is left unchallenged. Therefore, Cragg insists, many proponents of RJ make a mistake by seeking to remove the legal system from the sentencing process and replace it with informal methods of conflict resolution. A crime includes among its ‘victims’ the rule of law—indeed, it is often described as a “conflict” with the law—and part of the reconciliation has to be with the law. 53 Which is to say that he thinks proponents of RJ ought in large part to embrace the fourth criticism above that while the particular victim

has some rights, all citizens are victims in another sense and their rights need to be represented in the restoration of relations.

But this calls into question the view of sentencing as conflict resolution, since it is incompatible with the practice in most other forms of conflict resolution of having neutral mediation by a mediator accepted by both parties employing a number of widely known principles to which both sides can appeal, seeking to arrange a win/win situation. Since the judge is not permitted to be a neutral mediator between the offender and the law, how can the trial and sentencing be described as conflict resolution?\(^\text{54}\)

Cragg’s solution is to note that neutral mediation is merely a strategy for conflict resolution between equals, not a necessary feature of conflict resolution itself. If it is possible for the state to reduce its conflict with the criminal, then the state has the motivation to try. Building on Duff’s insight that trials seem to have some sort of communication as a goal, Cragg suggests that trials often do meet the commonly agreed upon practices of successful informal conflict resolution:

1) *Explain your own position and try to understand the other’s position.* The trial gives the state a chance to communicate what the law is, and if necessary, to state the justification of the law, while also giving the defendant a chance to introduce himself, explain relevant background information, and (with the aid of a defense attorney) to put his view of the case before the court.

2) *Try to defuse any initial hostility and to increase trust.* The legal system has no reason for hostility toward the defendant—indeed, this is its distinctive advantage over informal attempts at reconciliation with a party who feels wronged. Cragg writes that “there is no evidence that sentencing oriented toward putting defenders down, or teaching harsh lessons, or deterring through harsh sentences is effective.”\(^\text{55}\)

\(^{54}\) Cragg, *The Practice of Punishment*, p. 181.

\(^{55}\) Cragg, *The Practice of Punishment*, p. 182.
3) **Be friendly but firm, defending your legitimate interests while avoiding the temptation to inflict defeat.** A court can violate this by being too lenient or by being too strict. Both strategies are likely only to increase conflict with the law, and therefore cut against the purpose of the criminal justice system.

4) **Focus on common interests and propose solutions that the other person is likely to have an interest in accepting.** The common interest is clearly that the offender follow the law in the future. Cragg suggests that if attempting “to win the cooperation of offenders were not a high priority in modern systems, many popular sentencing options currently in use would be difficult to explain, including probation, parole, victim/offender reconciliation programs, community service orders, intermittent sentencing whose goal is to allow offenders to hold jobs and maintain community roles, and even rehabilitation.”

5) **Be prepared to make unilateral gestures of cooperation.** Cragg suggests that sentences that promise to help the offender (e.g., educative or rehabilitative programs), rather than sentences designed to teach the criminal a harsh lesson, can be described as such overtures.

Cragg thinks this view of sentencing makes it a central goal to persuade the offender to comply with the law voluntarily. This requires that the state try to understand why the criminals did not comply with the law in the first place, and to take those reasons into account when sentencing. Cragg argues that if the state keeps to the principle of using the minimum force necessary to protect those the law is supposed to protect, it will communicate to the criminal that the legal system is not out to get him, which is an important step towards winning his future compliance.

We’ve shown how Cragg can respond to most of the criticisms above, but we haven’t looked at the charge that the goal of conflict resolution leads to injustice because it requires indeterminate sentencing and different sentences for crimes identical in all


respects except in the ZIP code where they occur. There seems to be a sort of relativism inherent in the goal of conflict resolution, because the sentence is fitted not to the offence against justice, but to the victim and the community. It is consistent, therefore, with a community having unjust standards—it might very well be possible to resolve a conflict through an unjust solution so long as all parties agree to it. Cragg responds by claiming that in cases where there is widespread agreement about the evil of a particular crime (rape, e.g.), then there is usually substantial agreement about the severity of punishment. In cases where there is widespread disagreement (regarding abortion, e.g.), there is also disagreement about sentencing. The worry about “ZIP code justice” is thus a worry about pluralism in general. The objection makes the assumption that there is a widely held view of what is the just punishment in a given case. But that assumption is not always true—and where it is true, it is not clear that conflict resolution model won’t arrive at it or something close to it.

Cragg’s view can be seen to be a liberalism of fear that tries to minimize the use of violence, and to avoid it where possible. His view also has striking similarities to the Anselmian view we have been developing here, in large part because his goal of conflict resolution is so close to the Anselmian goal of right order. Cragg thinks that moral persuasion and education can occur in situations of unequal power relations, as does Anselm. Cragg thinks that justice requires a right will and not just right behavior, as does Anselm. Cragg’s recommendation that the courts approach their work with a conflict resolution mentality and using techniques that promote the resolution of conflicts seems

to give practical form to the Anselmian desire to show mercy to the recalcitrant. As with
Anselm’s view, Cragg emphasizes that the law has obligations to the entire public, which
places constraints on how flexible the legal system can be in its attempt to win the
compliance of the criminal.

Cragg’s view that the lawbreaker, simply by breaking the law, has committed a
moral wrong is analogous to Anselm’s view that presumption is the essence of injustice.
Cragg’s conception of the trial and sentencing process is similar to Duff’s, in that it
emphasizes its communicative and persuasive elements. But it is different from Duff’s,
and closer to the Anselmian view, in that it sees that the message communicated is not
that the particular law is morally justified, but simply that the offender is being treated
according to the law, that the verdict is justified by the law, that the court is bound by the
law, and that it is good that there be laws.\textsuperscript{60} Unlike Duff, Cragg does not hold that the
judge must “be of the view that the law being enforced is morally sound.” On Cragg’s
view, the judge must accept that he has “an obligation to enforce the law, independently
of his own assessment of its moral quality,” and he must be able to explain to an accused
why he accepts this obligation.\textsuperscript{61} The judge has the role of applying the law, not agreeing
with the law.

The Anselmian view we have been developing does diverge from Cragg’s view at
one key point. Cragg holds that violence is the \textit{summum malum}; that the use of violence,
which he sometimes describes as “the free use of force” is both immoral and anti-moral.

\textsuperscript{60} Cragg, \textit{The Practice of Punishment}, p. 232n23.

\textsuperscript{61} Cragg, \textit{The Practice of Punishment}, p. 232n24.
It is immoral, he claims, not only because of its threat to life and limb, but because “it attacks the capacity of human beings to live human lives” by disrupting the normal patterns of social cooperation at times and in ways that are unpredictable. Cragg thinks that “those harmed by the free use of force are unlikely to be harmed in ways consistent with any variety of justice.” Cragg thinks that violence is anti-moral because it impairs the ability of those affected by it to act as responsible moral agents. Those affected by violence are unable to plan their lives in accord with moral rules because violence disrupts the conditions for those plans. Violence forces those affected to become self-interested egoists rather than other regarding, which is an anti-moral perspective. Violence is often used to manipulate and coerce people, rather than to treat them as moral agents. For Cragg, then, punishment has something of the character of an evil, and his model sees the minimizing of coercion as minimizing an evil.

The Anselmian view does not hold the same view of coercive force. Such force has its place in a rightly ordered society, and it is just when used appropriately. The Anselmian view recognizes that force can sometimes be necessary to subdue the presumptuous and to protect the social order, in which cases the use of force is a duty of justice. On the Anselmian picture, disorder is the summum malum, that which is to be minimized; but sometimes minimizing the disorder requires the use of force. As we have noted numerous times, if people obey voluntarily there is more order than if they are coerced, since the goal is that people support due order for the sake of that order itself.

---


63 Cragg, The Practice of Punishment, pp. 88-89.
Therefore, authorities should punish only reluctantly and when it is necessary to do so.

But when it is necessary, coercion does not constitute an evil.

9.7 Indeterminate Sentencing and Social Defense

By placing such an emphasis on social order, the Anselmian view is committed to the view that a major concern of the state is the defense of that order from those in society who would disrupt it. Now, suppose a criminal refuses to repent, and instead expresses his intention to return to a life of crime. Wouldn’t the Anselmian view require that to preserve due order he be locked up forever, just as a recalcitrant sinner is condemned to hell to protect the right ordering of heaven?

There’s nothing in the Anselmian view to rule this out of bounds on principle. That this is true follows from its claim that human justice is continuous with divine justice, since divine justice allows for eternal punishment. This is actually an advantage for the Anselmian theory over those theories which do not permit perpetual punishment: since there are likely to be circumstances in which a regime will be required to restrain certain people indefinitely because of the extraordinary threat they represent, it is better that a theory of just punishment be able to accommodate these instances, explaining when such punishments are and are not just. There is a sense in which a theory of just punishment can be seen as a sort of domestic just war theory.

Does this mean that an Anselmian must dispense with determinate sentences, so that a petty thief arrested as a youth might as an octogenarian find himself still incarcerated? Not necessarily. Even in the cases where a criminal has rejected the offer of mercy and expressed his intention to return to a life of crime upon completion of his
sentence, we might nevertheless hesitate before deciding to “lock him up and throw away the key.” We might still release him if we thought he wasn’t likely to pose a direct and serious danger to the social order, or that the costs of incarceration were too high given the threat he posed, or that the surveillance that accompanies parole or probation would be sufficient to prevent him from carrying out his plan, or if we had reasons to doubt the strength of his intention to commit crimes in the future (we suspect his defiance of the law to be merely a boast before the other inmates). Perhaps, we think that we can communicate to him that we are releasing him not out of weakness, but because we think he might be able to go straight—evoking the classical use of mercy as a gesture of magnanimity to demonstrate the fundamental decency of the merciful person.

One might still object that these considerations are relevant to how one might exercise discretion within a framework of indeterminate sentences, but that they do not respond to the objections of those who think all sentences should be determinate. Is there a place for determinate sentences within this Anselmian view of punishment as social defense? I would think there could be, for reasons that overlap with the concerns of the liberalism of fear. A sentence of eternal punishment handed down by an omniscient, good, and just judge is one thing, but it is another thing entirely for such a sentence to be given by a relatively ignorant human being with the usual human vices and moral weaknesses. The corrections officers who deal with the inmates every day are the most likely representatives of the state to be able to make informed judgments about their charges. But a CO is under extraordinary stress, which might make it difficult for many of them to be objective assessors of an inmate’s attitude towards the law. Most of them don’t want to know too much about the inmates—not even the crime for which they were
convicted!—because of how it could affect their relationships with them, and yet such information would obviously be relevant for making decisions about their future.64

For the same reasons we want to separate the judicial and executive functions in society at large, we would probably want to separate the judicial and executive functions in making the decision about whether to release someone from prison or jail. Determinate sentences might be an efficient way of doing this, since they are less expensive than having separate parole boards or sentencing commissions. On the other hand, also for liberalism of fear reasons, we might want to avoid having the legislature determining sentences, since a get-tough-on-crime mood among the electorate might permit not just the cruel treatment of prisoners, but also the institutionalization of such treatment. Hence, constitutional safeguards against cruel and unusual punishments might make sense, as might the creation of sentencing commissions and parole boards that are somewhat insulated from quick changes in public opinion. (We might think it is important that society bear any costs necessary for the pursuit of justice.) These and other practical checks on the power to punish are justified, though, not because of a problem of punishment itself (as if punishment were an evil to be balanced against another evil), but because the practice of punishment can easily become merciless or otherwise inhumane as a result of human weaknesses and the pressures that crime can place on those responsible for order.

64 Ted Conover’s Newjack: Guarding Sing Sing (Random House, 2000) the best inside look at the work of a corrections officer, leaves no doubt that although a CO is “mainstream society’s last representative,” nevertheless “it is a long time after a new officer enters upon his duty before he becomes, even under the most favorable circumstances, fully competent to discharge it,” if he ever does (quotes from pp. 207-8).
With these caveats and cautions in mind, though, the Anselmian view could in principle support two controversial practices associated with societal defense: capital punishment and the recent trend of “Sexually Violent Predators” laws which incapacitate certain criminals after their sentence has been served because it is determined that they are a continuing threat.

Proponents of the death penalty often justify it as appropriate retribution for particularly heinous crimes, especially homicide and treason, or as a necessary deterrence against such crimes. Since the Anselmian view rejects both retribution and deterrence as appropriate motives for punishing, the only justification of capital punishment available to it would be societal defense. If a criminal poses or is likely to pose a danger to others even if incarcerated, then it would be necessary to execute him or her. If incarceration would be sufficient to protect society, then that would be sufficient. It might very well be the case, as Pope John Paul II suggests, that the number of cases in modern societies where only the death penalty would protect society is “exceedingly rare, if not practically non-existent”. But one could argue that in cases where a prisoner commits a violent crime while incarcerated, the available methods of incarceration have shown themselves to be inadequate to defend other human beings against him, so that the death penalty would be necessary. In this case, each instance of violent crime within a prison ought to trigger a reexamination of how the facility is run, to see if there is some way of avoiding


a repetition of the crime. However, since it seems impossible to have a reasonably humane prison that eliminates every chance of a prisoner committing a violent crime, it very well may be true that there will always be some instances where the defense of the other prisoners requires that the violent prisoner be executed.

A similar sort of calculation could justify the various laws that allow states to keep some criminals in custody after they serve their sentence if it is determined that they are likely to commit certain violent crimes were they to be released. In 2003 the Virginia legislature passed its “Sexually Violent Predators Act” which contains the following provisions: As the term of a prisoner convicted of a sexually violent crime nears completion, a review committee screens, evaluates, and makes recommendations to the attorney general’s office whether the prisoner be released outright, be placed in a conditional release program, or be civilly committed to a secure mental health facility as a sexually violent predator. In the last case, the offender would have his case reviewed by the court periodically to determine whether or not he remains “a sexually violent predator.” Even with frequent reviews, it is entirely possible that he might never be released.\textsuperscript{67} Virginia’s Attorney General Jerry Kilgore gave the following justification for these laws\textsuperscript{68}:

“When we talk about sexually violent predators, we are not talking about other types of criminals who may age out of their crime-prone years after a long stretch in prison. We are talking about


\textsuperscript{68} Ibid.
individuals who will – if left free – commit their crimes over and over again.”

Kilgore’s justification has nothing to do with deterrence or retribution. Rather it is an argument that these criminals should be incapacitated in the interests of public defense. The Supreme Court found in *Kansas v. Hendricks* (1997) that so long as the criminals were not held in prison indefinitely, their being committed to a secure hospital after a civil proceeding did not constitute punishment and their post-sentence review did not count as double jeopardy.69

Proponents of retributivist theories of punishment could not endorse such laws, obviously. Nor is it likely that the law could be justified on deterrence grounds—not only are the laws not designed for the purpose of deterring would-be criminals from the crime, but one would also expect the deterrent effects of such laws to be rather minimal, since there would be a great degree of uncertainty about whether a would-be criminal would be moved to a hospital at all.

For the Anselmian view we have been developing, though, there is nothing in principle wrong with these laws. As we argued above, there might be serious objections from the perspective of the liberalism of fear that indeterminate sentencing puts too much discretion in the hands of the state. However, the liberalism of fear does not have a principled objection to these laws, so that such fears could be allayed if the statute has

69 *Kansas v. Hendricks* (1997), 117 S. Ct. 2072: “The Act does not establish criminal proceedings, and involuntary confinement under it is not punishment. The categorization of a particular proceeding as civil or criminal is a question of statutory construction.”
sufficient safeguards against abuses. Arguably, it is an advantage of the Anselmian view that it allows the state to act in the defense of its citizens in this way.
CONCLUSION

To our knowledge, nobody else has tried to draw a connection between Rawls and Anselm before, let alone to argue that Anselm’s views on justice can serve as a corrective to those of Rawls. The key to our argument here has been first to note that Rawls’s arguments for stability and autonomy pull apart, then to show that stability is the more basic and more neutral of the two ideas, and finally to show that Rawls’s notion of stability dovetails nicely with Anselm’s account of rectitude. Only then could we argue that Anselm’s view of justice and mercy as based in rectitude provides a more stable view of politics than Rawls’s justice as fairness.

We have argued that Rawls’s account of justice is unstable because of its commitment to using the coercive power of the state to promote a regime based around a view of autonomy that not all citizens have reason to accept. We did not argue that Rawls’s view of autonomy is immoral or defective in itself, but only that it cannot provide the basis for a stable account of liberal justice according to Rawls’s own principles. Under the regime organized around Anselmian principles, citizens who share Rawls’s view of justice as fairness are free to promote their view among their fellow citizens, and if they can reach a stable political arrangement based on such principles while remaining within the constraints of mercy, that is their prerogative. Likewise, citizens who want to promote other arrangements, perhaps even arrangements looking like feudalism, are free to do so, so long as they respect the demands of mercy. (Of
course, the more eccentric the political arrangements one proposes, the less likely people are to find a principled reason to support it, and so the more heavily the demands of mercy will be obstacles to those arrangements being realized.

We have shown that an Anselmian account of justice and mercy, precisely because it is an account of justice and mercy provides us with an angle from which to develop a view of punishment that has serious advantages over the major accounts of punishment available, is consistent with current practices, and provides guidelines for reform.

Finally, we have argued that Anselm’s theory provides a significant and powerful account of justice, mercy, order, and freedom, which has been overlooked or misunderstood for centuries. Anselm’s view uses feudal terminology to articulate certain concepts, but his account of justice and rectitude are not tied to feudalism—indeed, we have argued that they are applicable to morality and politics in the twenty-first century. We have argued that Anselm has an account of concepts and language regarding the divine attributes that is different from the more familiar (non-apophatic) accounts based on analogy. Anselm applies this theory to bring human and divine notions of justice into a common intellectual framework, and even seems to have appealed to his theory in shaping his policies and agendas as Archbishop of Canterbury.

Given all this, we might ask ourselves what Anselm would think of our “Anselmian” Liberalism of Mercy? Would he recognize it as something derived from his thought? Would he approve of its account of punishment? Would he reject its adaptation to liberal democracies? Obviously, we cannot know the answer, but we can make some educated guesses. Liberalism is designed to be frustrating to those who have clear ideas
about how the world can be, because liberalism requires not only that we answer our opponent’s objections to our plans in ways that satisfy us, but also that we do our best to win our opponent over to our side. Anselm’s tenure as archbishop brought him face to face with such frustration, as first William II and then Henry I prevented him from exercising his office in the way that he thought necessary. His entire life from the time he was elected prior of the abbey at Bec made clear to him just how difficult it was to win the obedience of prideful men, but also how powerful a tool was mercy in doing just that. If Anselm found himself in a position where his substantive views of justice were in the minority, he would obviously be frustrated. But if the regime were designed around his formal view of justice as rectitude, and if it were to make a serious attempt to encourage mercy among its citizens and to build mercy into its view of the law, then he just might recognize it as his own. In any case, he would certainly roll up his sleeves and get to work trying, as he once put it, “to repress evils by pastoral rule, to coerce the presumptuous, and to lead everything irregular back to due order.”

1 Aep 198.

**WORKS OF ANSELM**


**OTHER LITERATURE ON CLASSICAL AND MEDIEVAL PHILOSOPHY**


--------. “Re-reading De Grammatico or Anselm’s introduction to Aristotle’s Categories.”


Kempshall, M. S. *The Common Good in Late Medieval Political Thought*. Oxford University Press, 1999.


**WORKS OF JOHN RAWLS**


**SECONDARY WORKS ON RAWLS AND LIBERALISM**


Daniels, Norman. “Reflective Equilibrium and Justice as Political.” In Davion and Wolfe, 127-54.


Friedman, Marilyn. “John Rawls and the Political Coercion of Unreasonable People,” in Davion and Wolf, 16-33.


-------. “Civic Education in the Liberal State.” In Rosenblum, Liberalism and the Moral Life.


-------. *The View from Nowhere*. Oxford University Press, 1986.


--------. “Contractualism and Utilitarianism” in Sen and Williams, 103-128.


Wolterstorff, Nicholas. *John Locke and the Ethics of Belief* (Cambridge University Press, 1996);


WORKS ON THE PHILOSOPHY OF PUNISHMENT AND MERCY


--------. “The Responsibility of Criminals.” In Acton, The Philosophy of Punishment, 175-183


Mabbot, J. D. “Punishment.” In Acton, 39-54.


Murphy, Jeffrie G. “Marxism and Retribution.” In Duff and Garland, 47-70.


Quinton, Anthony M. “On Punishment.” In Acton, 55-64.


