

Liberal Politics and Public Faith

A Philosophical Reconciliation

Dissertation Submitted to the University of Arizona
In Partial Fulfillment for the Award of Doctor of Philosophy
In Philosophy

Kevin Vallier

Under the Guidance of Gerald Gaus



University of Arizona
Tucson, Arizona
March 2011

Contents

	Pages
<u>Introduction</u>	1
<u>Part I: Conflict</u>	11
<u>Chapter 1: The Liberal Idea</u>	12
<u>Chapter 2: Objections of Faith</u>	34
<u>Chapter 3: A Model of Liberal Practical Reason</u>	70
<u>Chapter 4: A New Model of Liberal Practical Reason</u>	97
<u>Part II: Reconciliation</u>	138
<u>Introduction: Two Metrics and Two Variables</u>	139
<u>Chapter 5: Conceptions of Public Reasons</u>	143
<u>Chapter 6: Against the Standard Conception of Public Reasons</u>	162
<u>Chapter 7: Convergence</u>	189
<u>Chapter 8: Conceptions of Idealization</u>	211
<u>Chapter 9: Against the Standard Conception of Idealization</u>	235
<u>Chapter 10: Moderate Idealization</u>	253
<u>Conclusion</u>	285
<u>References</u>	291

Introduction

The wedding of liberalism and secularism has alienated millions of Americans from the liberal tradition. To many, liberalism is identified with an aggressive secularization of social institutions, especially public media and public schools. Many citizens of faith believe liberalism threatens their most sacred forms of life. This was not always the case: in American history, the relation between liberalism and religion has often been one of mutual respect and support. Mutual respect has given way to mutual suspicion. Liberalism's proponents *and* detractors believe that liberal political theory requires eliminating religious influences in public life. This dissertation defends a bold thesis: *liberalism has no such implications*. One we see why this is so, the philosophical underpinnings of the liberalism-religion conflict are undermined. Unlike most work on this topic, the dissertation neither defends liberal secularism nor assaults it. Instead, it integrates the moral emphases of both the religious and liberal perspectives.

And yet *so much* has been written on the clash between faith and liberal democratic politics. What could possibly justify another article, much less an entire dissertation? My justification for adding yet another work to the pile is that I approach the liberalism and faith conflict from a new angle—I attempt to carefully articulate the philosophical structure of the conflict and resolve it without taking sides. What has been written heretofore either defends liberal secularism or assaults it (if to varying degrees). I propose to make an end-run around the debate, to attack its shared premises and carve out a position that integrates the moral emphases of both camps while chastening them for being one-sided.

But before I begin, I must articulate this conflict with great care. There is no one religion and politics problem. There are many religions and many forms of political life. Neither politics nor religion is a univocal phenomenon. Accordingly, there will be no one way in which they clash. Our focus will be on the role of religious expression and religiously-motivated political action in *liberal democratic* politics and the *liberal political tradition*. This is the problem as we know it in the United States and Western Europe. In fact, to be perfectly frank, the problem of religion in public life in the United States and Western Europe is a problem with certain forms of Christianity, Judaism and Islam. In our ideological neck of the woods, Hinduism, Buddhism, Confucianism, Taoism, Shintoism, and Janism simply do not raise these issues. Accordingly, I focus on forms of the major monotheisms that demand great devotion to the dictates of God and that require their adherents to

shape their social and political world in accord with religious values. Within Western democracies, these religious have millions of adherents and accordingly they raise serious practical conflicts.¹

We can focus further still. This work is situated within a debate between (mainly) secular public reason liberals and their (principally) Christian critics. It is for this reason that I will focus primarily on the problems raised by Christianity in public life. As of this writing, I do not have a worked out view about Islam, though I suspect much of what I say here will have important implications for the conflicts between Islam and liberal democratic political cultures. In particular, I imagine that worries about religious integrity will be more pressing for many Islamic communities. Hundreds of millions of Christians have lived under liberal democratic institutions over the course of hundred years, sharply contrasting with the relatively recent encounter between certain forms of Islam and Western liberal democracies. Working out their mutual compatibility is a young project for Muslims relative to the Christians.

And yet, in another way, this work is not about religion alone. Instead, its deepest concern is with the problems posed by the *practical demands* of these faiths, demands which may well arise for any number of non-religious doctrines, ideologies, projects and social practices. I am concerned about forms of life that require *social integration* or that make social and political demands of their adherents. I am worried about citizens who believe that their individual ideals, ideals not shared by their compatriots, should shape and structure their lives regardless of the views of others. Our problem then is in many ways the problem of the social contract tradition as a whole: how are divergent forms of life supposed to coexist in accord with a public social morality that allows all citizens to live well together? John Rawls has asked this question powerfully:

How can it be reasonable or rational, when basic matters are at stake, for citizens to appeal only to a public conception of justice and not to the whole truth as they see it? Surely, the most fundamental questions should be settled by appealing to the most important truths, yet these may far transcend public reason!²

¹ I set aside discussions of the roles of Eastern religions in democracies like India for the sake of focus. However, I must note that Judaism raises important problems of its own, though they are of distinctive and unusual varieties, such as Judaism's role in Israeli politics and in the relationship between Israel and other liberal democracies. It is unclear to me whether as a result Judaism is at all incompatible with liberal democratic public life. My sense is that in American political life, followers of reformed Judaism like their liberal Protestant counterparts are largely unconcerned about public secularization imposed by the liberal political tradition. Similarly, conservative and orthodox sects of Judaism will likely have similar concerns about the restrictions on religious integrity imposed by the mainstream interpretation of liberalism as do their conservative Catholic and Protestant counterparts.

² Rawls 2005, p. 216.

Rawls's question is our question too. How can it be rational for citizens of deep conviction (not merely people of faith) to abide by social rules that require them not to act on their commitments? Clearly we need a set of social rules that promote social cooperation and we know that these rules will frequently require citizens of conviction to restrain their social and political action. A well-ordered society, one whose institutions treat all with dignity and respect, requires that citizens compromise with one another. Truly liberal citizens accord one another the freedom to live their lives as they see fit, yet towards this end they ask one another to refrain from institutionalizing their competing value systems. In short, liberalism respects freedom by restricting it. Consequently, liberals and citizens of faith (groups that may well overlap) see the other as attempting to restrict freedom. The person of faith sees liberalism as oppressive because it asks her to "split" her public and private selves, or to privatize her beliefs, and the liberal sees the person of faith as oppressive precisely because she refuses to do so. Liberalism and religion appear to place competing demands on citizens. But liberalism wants a society where all are free. I shall argue that the only way to achieve such a society is for liberals to allow all citizens of conviction to engage in political action based on their deepest convictions, while remaining true to the liberal idea.

We must be more specific still, for just as there are many religions, there are many *liberalisms*. I cannot hope to resolve the conflict between religious convictions present in liberal democratic societies and all liberalisms. Instead, I will focus on the academically dominant form of liberalism and the form of liberalism I regard as the most developed and attractive, what is known as *public reason liberalism*. Public reason liberalism combines the traditional liberal commitments to basic liberal democratic liberties (freedom of religion, speech and press) and political equality (universal suffrage and democratic governance) with the demand that coercion and political authority be *publicly justified*. Thus public reason liberalism holds that for the coercive power and authority of the state to be legitimate and just, it must be justified in terms that each person can reasonably be expected to accept. Public reason liberalism requires that coercion and authority be justified to all *by their own lights*. Public reason liberals frequently argue that respect for persons as free and equal requires that coercion be publicly justified, for unless John's coercing Reba can be justified to Reba on her own terms, John simply imposes his will on Reba, his moral and political equal. Reba is a naturally free person and, as such, for John to have the right to force Reba to change her behavior, the right in question must be one to which Reba is rationally committed.

Within the world of political philosophers, political theorists, political scientists, legal scholars, historians, sociologists and public intellectuals, public reason liberalism takes center stage. While other forms of liberalism are frequently developed and defended in other literatures, the religion and political literature has tended to focus on this single target for both assault and defense. In this I will follow the literature. I do so not merely because “everyone else is doing it” but because the literature is focused in this way for a reason. Public reason liberalism unlike alternative liberalisms like liberal perfectionism is concerned about the justification of coercion to others in their own terms and so it claims for itself a kind of impartiality that is both attractive and potentially suspect. In other words, it raises a unique and compelling philosophical problem. In the end, I side with the public reason liberals with respect to their fundamental commitment to the public justification of coercion, but I do so only after arguing at length that public reason liberalism must fundamentally change in light of the criticisms lodged against it.

Who issues these criticisms? They typically divide into two classes. We might call the first class “The New Traditionalists,” philosophers and theologians like Alasdair MacIntyre, John Milbank, and Stanley Hauerwas who reject liberalism because it is incompatible with religious, or more generally virtuous, transcendent and communal forms of life.³ The second class of philosophers and theologians, such as Kent Greenawalt, Michael Perry and Nicholas Wolterstorff, endorse the basic institutions of liberal democracy along with its associated rights and duties despite holding that public reason liberalism places undue strain on persons of faith.⁴ Discussion focusing on this latter group promises to be more productive because of their shared commitments with public reason liberals. I call this second group of critics *integralist* liberals for while they endorse liberal democracy they typically maintain that public reason liberalism prevents citizens of faith from living religiously integrated lives.⁵ This work therefore takes its place within a philosophical dialectic between those committed to the broad outlines of liberal democratic political theory. It does not raise the question of whether liberal democratic institutions are justified *as such*. Instead, its audience is those already committed to these institutions but who profoundly and fundamentally disagree about the place of religious conviction within them.⁶

³ See Hauerwas 1999, pp. 13-22, MacIntyre 1990, pp. 170-195, Milbank 2006, pp. 9-50 for some of the New Traditionalists’ concerns in their classic works.

⁴ These theorists include Christopher Eberle 2002, Kent Greenawalt 1995, Michael Perry 1991, Philip Quinn 1997 and Nicholas Wolterstorff 1997.

⁵ I have borrowed this useful term from Benjamin Hertzberg.

⁶ We might follow John Tomasi as describing these two groups as defenders of “A-people” and “C-people” though describing one side of the debate as “As” and the other as “Cs” is arguably prejudicial. See Tomasi 2001, pp. 17-26.

This work approaches the criticisms of integralists from *within public reason*. Integralists have distinct religious, political and philosophical commitments and as such their criticisms are frequently motivated by concerns that will not move public reason liberals. If we approach integralist criticisms as embedded in their own conceptual frameworks, the task of assessment will be overwhelming. Instead, my aim will be to show that public reason liberals must address the integralist criticisms from within *their own theoretical commitments*. This is not to say that this work does not speak to integralist liberals. Instead, in developing a version of public reason liberalism that answers the integralist, I seek to bring the integralist closer to the public reason liberal tradition. While I do not provide arguments that integralists should *be* public reason liberals, in disarming their objections I hope to make public reason liberalism more attractive to them.

I hope that my integralist interlocutors will see my work as a good faith attempt to take their concerns seriously. Public reason liberals often deride integralists. One notorious example is Stephen Macedo, who claims that “If some people ... feel ‘silenced’ or ‘marginalized’ by the fact that some of us believe that it is wrong to seek to shape basic liberties on the basis of religious or metaphysical claims, I can only say ‘grow up!’⁷ I aim to treat integralists in precisely the opposite way. I believe their concerns are so serious that they can only be adequately addressed by conceptual work within normative ethics. This work is an exercise in mutual respect, not only in its aim of developing a version of public reason liberalism that integrates the concerns of both sides, but that honors both sides in its argumentative structure and prose. I hope in this way to invite integralist and the public reason liberal to view each other with mutual respect.

I believe that the conflict between public reason liberals and integralist liberals is fundamentally a conflict between integrity and treating others with respect by reasoning from their standpoint. Both sides have a *liberal* concern: preserving the freedom of all citizens to live by their own lights. The integralist emphasizes the authority of the individual to live an integrated life, a life where her public and private endeavors are shaped and informed by her convictions. The public reason liberal emphasizes the need for citizens to reason from each other’s standpoints in order to jointly construct a social morality that allows us all to cooperate while simultaneously living in accord with their distinctive values. From the perspective of liberalism, *how could either side be wrong?* Surely the interesting question for the liberal is not “Who is right?” but rather “Both sides have a point. What now?” The latter question is my question. I believe that an answer can only be found at the conceptual level. Here is my statement of **The Problem**:

⁷ Macedo 2000, p. 35.

There are two fundamental sources of normativity, a first-personal source and a second-personal source, that conflict in a complex manner. Each side in the debate draws from one of these two sources. The public reason liberal draws on a second-personal source of normativity by emphasizing the need to reason from the standpoint of others, while the integralist draws on a first-personal source of normativity, the authority of the individual to live an integrated life. And yet a viable liberal political theory must integrate these two sources of normativity. How is this to be done?

I solve **The Problem** by accomplishing five aims:

Aim 1: State precisely what the dispute *is* and why it is philosophically interesting.

Aim 2: Validate the concerns of integralists while protecting public reason.

Aim 3: Show that public reason liberalism must change in light of integralist criticism.

Aim 4: Provide a strategy for discerning how the change must proceed.

Aim 5: Execute the strategy by defending an alternative version of public reason liberalism.

Achieving each aim is a great endeavor in itself. Aim 1 is particularly challenging. Many people argue about the problem of the proper place of religion in the public square without clearly stating it. I spend *four chapters* trying to accomplish this aim, while simultaneously achieving Aims 2, 3 and 4.

I must say something here about my approach. Participants in this literature, be they academic or public pundit, engage it from a myriad of perspectives and methods. My method is that of a contemporary analytic political philosopher. My skill set lies primarily in conceptual analysis and the assessment of arguments at a high degree of abstraction. For this reason I will not take on the sociological concerns of many public reason liberals, such as claims that a desecularized public square will produce bad consequences, like religious violence.⁸ What I am after are arguments about the clash between first and second-personal sources of normativity. That is, I am interested in claims like the following: does the conceptual structure of public reason liberalism require restraints on religious expression in the public sphere? Does respect for persons require the public sphere to be secularized? Can the account of justificatory reasons offered by mainstream public reason liberalism withstand sustained philosophical scrutiny? Some will find these questions uninteresting and esoteric, but I claim that if **The Problem** can be resolved at the conceptual level that great social

⁸ Christopher Eberle addresses some of these concerns, see Eberle 2002, pp. 152-186.

and political fruit can be borne. Not only will we have resolved an important philosophical dilemma, but the new conceptual space uncovered by my inquiry will have important implications for the structure of constitutional law, public policy and civility within political discourse. I cannot convince you of this now, so I write a promissory note: by the end of the dissertation, I will show that the conceptual work accomplished herein will matter outside of philosophy.⁹

We may now turn to the structure of this work. It contains ten chapters divided into two parts. In Part I, I outline the structure of public reason liberalism and analyze the standard integralist objections leveled against it. I argue, in short, that the conflict between public reason liberals and their integralist critics is an acute version of a conflict that lies at the heart of social morality. In short, it is a clash between individual integrity and the need for interpersonally justified moral rules. Chapters 1 and 2 explain the normative concerns of both sides and the structure of public reason liberalism. They depict the tension between public reason liberals and integralists.

To elucidate this tension, I draw on the resources of contractualist moral theory. After criticizing mainstream contractualist views like those of Darwall and Scanlon in Chapter 3, I develop an alternative conception of contractualism that employs two types of normative reasons—first and second-personal reasons—in Chapter 4. First-personal reasons are reasons to act in accord with our most important projects, plans, principles and ideals—that is, to live lives of integrity. Second-personal reasons are reasons to conform one’s behavior to rules that can be justified from the standpoint of others. I argue that while social morality includes both types of reasons, public reason liberals tend to emphasize the second-personal while their critics emphasize the first-personal. Public reason liberals believe that a second-personally justified political morality requires not advancing unshared considerations in public political life; they thus advocate restraint on religiously-motivated political behavior. In contrast, critics emphasize religious believers’ duties to shape their institutions according to their principles and values. Restraint should be rejected, these critics argue, because it demands that religious citizens “privatize” their beliefs. A complete political theory must grapple with the inevitable tension between the two. I complete Part I by arguing that a conception of public reason commends itself to the extent that it can accommodate and resolve the tension.

⁹ And yet, the reader will notice that my method is not limited to those of the analytic political philosopher. Instead, I reach out more broadly for tools from normative ethics, metaethics, and epistemology, while in future versions of this work I will also reach out to economics, legal theory, political science and public policy for help applying the theory developed here to real public controversies.

In Part II, I employ the framework developed in Part I to reconcile the ideal of public justification and the integrity-based claims of citizens of faith. Reconciliation requires developing an alternative conception of public reason liberalism that imposes few integrity costs. The alternative conception contains two elements, a *convergence conception of public reasons* (developed in Chapters 5, 6 and 7) and a *moderate conception of idealization* (developed in Chapters 8, 9 and 10). The convergence conception of public reasons holds that reasons offered to justify coercion need not be shared; a convergence conception of reasons will therefore be unable to motivate restraint. Another key aspect of the debate is the extent to which we idealize the subjects of justification, i.e., how far we assume that they have adequate or complete information, are excellent at reasoning, and so on. The orthodox conception of public reason embraces radical idealization, or imputing full information and perfect rationality to subjects of justification. Such idealization, however, abstracts citizens from their conceptions of the good and threatens to violate their integrity. In contrast, I defend a more moderate conception of idealization which attributes reasons to citizens on the basis of their deeply held principles and values while correcting for gross errors in reasoning and obvious informational deficits. While the religious objections provide reason to reject orthodox public reason liberalism, I conclude that a version of public reason liberalism that embraces convergence and moderate idealization can meet the criticisms of integralists.

Acknowledgements are, of course, in order. I must begin by thanking three individuals who prompted my thinking on this topic years ago. In the fall of 2004, Gregg Keithley challenged me to wonder how religious toleration could be justified and, in particular, why some religious sects should not persecute others in the name of religious truth. I must also thank Emily and Mark Laufgraben for helping me see the gravity of the questions Gregg pressed me to ask. Without them, I may never have embarked on this dissertation project.

When I began my graduate work at the University of Arizona in the fall of 2005, I let my questions about the role of religion in the public square die down so that I might pursue other subjects. But the questions simply would not go away. In 2006, Jerry Gaus encouraged me to address these questions once more. It was Jerry who helped me to see that an Aristotelian-Thomistic approach to these matters was inadequate and that only an approach to social and political theory informed by the social contract theorists—Rousseau, Kant and Rawls in particular—could succeed in quieting my concerns. Jerry eventually became my adviser and has put more work into my academic development, and this dissertation, than anyone could reasonably hope for. For that I am deeply grateful.

In my early writing on these matters, I have many people to thank. But most importantly, I must thank Chris Freiman and Nathan Ballantyne for many conversations on the matter and for the great effort they put into my writing. Bill Glod contributed much to these discussions and edits as well. Along the way, I presented my work at a number of conferences and seminars, including the Arizona Dissertation Seminar, several conferences sponsored by the Institute for Humane Studies, graduate conferences at the University of Arizona, Arizona State University, St. Louis University, the City University of New York and the 2010 meeting of the American Political Science Association. In the process I received very helpful feedback from many people, including Thomas Christiano, David Schmidtz, Steven Wall, Uriah Kriegel, Ian Evans, Brian Fiala, Helen Daly, Stephen Lenhart, Robbie Wagoner, Rachana Kamtekar, Benji Kozuch, Michael Bukoski, Javier Hidalgo, Jason Walker, Steven Davies, Bas van der Vossen, Benjamin Hertzberg, Stephen Macedo, Paul Weithman, Simone Chambers and many others. I want to thank my friends Kyle Swan, Mark LeBar, Justin Tosi and Enrico Zaffoli for working through earlier drafts. Keith Hankins, Chad Van Schoelandt and John Thrasher worked especially hard to help me improve the penultimate draft. Chad was kind enough to proofread the dissertation. And I must mark John out for special thanks for many long, heated arguments about the nature of religious belief and its place in politics. John continuously pushed me to recognize the enormity and severity of these issues.

While I have thanked the internal members of my dissertation committee—Jerry Gaus, David Schmidtz, Thomas Christiano and Steven Wall—I must go out of my way to thank Christopher Eberle and Nicholas Wolterstorff for sitting on my committee as external members. They are two of the most prominent and insightful integralist critics, and so my dissertation was immensely improved by the hard work they put into reading it. I am deeply honored by their investment and cannot thank them enough.

Finally, and most of all, I must thank my wife, Alicia, for the emotional support required to write a long dissertation over the last three years. Whenever I would become discouraged by an argument or criticism, she was there to push me forward. Whenever I thought the dissertation was without merit, she assured me that it was. And whenever I was ready for a break to do something fun, she would always join me. No one deserves such kindness, strength and wisdom combined into a single person.

Much of the work herein has been published in article format and has benefited from the refereeing process. A version of Chapter 6 will be published in *The Journal of Moral Philosophy*, “Against Public Reason’s Accessibility Requirement”, so my thanks goes to the JMP’s editor, Thom

Brooks, a version of Chapter 7 will be published in *Public Affairs Quarterly*, “Consensus and Convergence in Public Reason”, so I thank that PAQ’s editor, Bob Talisse, and a version of Chapter 2 will be published in *The Australasian Journal of Philosophy*, “Liberalism, Religion and Integrity”, so I thank the AJP’s editor, Steward Candlish. I also greatly appreciate the comments of anonymous referees for my pieces at the *JMP* and the *AJP*.

With this, let us turn to Chapter 1, where I will lay out the structure of public reason liberalism and the moral ideal that motivates it.

Part I

Conflict

Chapter 1: The Liberal Idea

This work's thesis is that public reason liberalism, despite the doubts of many, allows religion to play a major role in politics. Since public reason liberalism is the dissertation's primary subject, I begin by explaining public reason liberal political theory. To put it briefly, public reason liberalism is the branch of liberal political theory that requires coercive impositions be justified by an appeal to public reason. Coercion, whether authorized by a law or policy, must be justified to those coerced in terms they can reasonably be expected to endorse.

This chapter divides into five sections. In Section I, I will cover the core motivation behind public reason liberal political theory: appeals to truth claims alone cannot justify coercing others. To illustrate, I will focus on one of Locke's defenses of religion toleration. In Section II, I will explain liberalism's presumption against coercion, which holds that each person is by nature free and equal and that as such there is a presumption in favor of her being able to direct her life as she sees fit. I will then tie the presumption against coercion to the ideas of public reason and public justification. Section III articulates and explains the Public Justification Principle, the defining principle of public reason. Section IV outlines two of its central aspects that take center stage in future chapters. Section V concludes. The contrast between public reason liberals and their religion-friendly critics is their contrasting emphases on the normative force of the standpoint of others and the standpoint of individual integrity. This chapter explores the connection between public reason and taking the standpoint of others. In doing so, it describes the principles embraced by one side of the debate, generating the first half of our description of the problem of the proper role of religion in politics.

I: The Basic Idea

It may be unfair to state, as Perez Zagorin does, that “[o]f all the great world religions, past and present, Christianity has been by far the most intolerant.”¹ Nevertheless, the claim highlights a significant historical fact: until the twentieth century, fights among Christians have often been bloody and intractable. For much of the medieval period, Roman Catholicism was not only religiously dominant, but the focal point of European social and political life. During the Reformation, the Roman Catholic Church's social and political hegemony broke down. After

¹ Zagorin 2003, p. 1.

decades of unusually violent conflict, previously Roman Catholic Europe transformed into the modern nation-state system. But even until the eighteenth century, stable cooperation between distinct Christian sects was rare. Many political traditions struggled with this problem of social unrest, a struggle only resolved, paradoxically, through the conflict over religious toleration. Religious toleration, peacefully coexisting with those one significantly disagrees with, came to Europe slowly and painfully.

Liberalism was the first political tradition to defend religious toleration as a political solution to religious war and disagreement. Yet its emphasis on toleration originated not only in a desire for peace. Many religious political forces maintained that a stable political order was impossible when citizens disagreed about fundamental matters. Much of liberalism seeks to show that political order can be achieved despite wide differences in beliefs, even on matters as significant as religion. While Thomas Hobbes and John Locke supported attempts to stamp out Catholicism in England, they rank among the earliest figures to argue that social cooperation could be achieved by appealing to the common reason of all instead of the truth of a particular religious tradition.² And while a number of religious and political thinkers before them believed that political justification could be achieved by natural reason, these earlier theorists did not struggle with the problem of persistent and profound disagreement over religious matters that became acute during the seventeenth century.³

Religious toleration nonetheless prevailed in Western societies. In the twenty-first century, these countries still practice religious toleration extensively, if not universally. We cannot delve into the historical reasons the idea of religious toleration took hold. But the philosophical challenges associated with religious toleration go to the core of the liberal tradition. The “Liberal Idea” that results from these challenges holds that political coercion must be justified from the standpoint of others. To explicate the Liberal Idea, let us turn to one of Locke’s famous arguments for toleration, an argument that contains the Liberal Idea in its infancy.⁴

² For standard examples of the “intolerance” of social contract liberals, see Hobbes 1994, pp. 411-488, Locke 2003, p. 245, Rousseau 1997, p. 151.

³ There is a notable absence of concern with political implications of disagreement throughout the great works of political philosophers even into the 17th century. Pierre Bayle and Roger Williams stand out as rare exceptions. See Bayle 2000, p. 235, Williams and Davis 2008.

⁴ Some will disagree with my exegesis, yet even if the exegesis is wrong, the interpretation I defend still illuminates the Liberal Idea.

I.ii Locke

Locke presents his case for religious toleration in *A Letter Concerning Toleration*. Among his arguments are claims that intolerance leads to bad consequences, that it is bad for personal virtue, that it is intrinsically unfair and that goes against God's will.⁵ Many of Locke's arguments depend on empirical considerations and complex theological points, but we will focus on his argument from reciprocity:

But if one of these churches hath this power of treating the other ill, I ask which of them it is to whom that power belongs, and by what right? It will be answered, undoubtedly, that it is the orthodox church which has the right of authority over the erroneous or heretical. This is, in great and specious words, to say just nothing at all. *For every church is orthodox to itself*; to others, erroneous or heretical. For whatsoever any church believes, it believes to be true and the contrary unto those things it pronounces to be error. So that the controversy between these churches about the truth of their doctrines and the purity of their worship is on both sides equal.⁶

Locke speaks of a church having the “power of treating the other ill” and asks how such power can be justified. I take it that he is referring to a power of *authority* in that the “orthodox” church has the power to *obligate* unorthodox churches to obey its commands in matters of doctrine and worship.⁷ The argument is rooted in Locke's analysis of the kinds of claim that natural equals can make obligatory for one another.

Locke may initially appear to argue from fallibilism. While every church believes it is orthodox, no church can know with certainty that it is right and so a given church has no authority over non-members because it might be wrong. I deny this: Locke is instead claiming that a church cannot acquire authority over non-members because its authority cannot be justified by appeal to its orthodoxy *regardless of whether it is orthodox*. A claim to orthodoxy cannot ground authority even if true.

⁵ See Locke 2003, pp. 215-218 where Locke questions the virtue of persecutors, for some theological arguments for toleration, see *ibid.*, p. 219. Locke argues that giving the church the authority to enforce orthodoxy has led to bad consequences on *ibid.*, p. 237.

⁶ *Ibid.*, p. 225, emphasis mine.

⁷ To simplify, I will assume that a person who is “orthodox” has the true view about the proper interpretation of the relevant religion.

We should be clear about the particular view Locke is arguing against. It might be any one of the following three claims:

- (i) That John is *in fact* orthodox gives him the authority to enforce his theological views.
- (ii) That John is *epistemically justified in believing he is* orthodox gives him the authority to enforce his theological views.
- (iii) That John is *justifiably recognized by all relevant parties as* orthodox gives him the authority to enforce his theological views.

In support of (i), Locke argues that a church's assertion that it is orthodox is tantamount to "say[ing] just nothing at all." This is to say that whether John is orthodox is irrelevant to determining whether John has the authority to demand that others change their practices to meet John's standards. Consequently, that John is orthodox cannot add to John's authority. If the fallibilist interpretation were correct, Locke would not argue a correct assertion of orthodoxy somehow fails to ground claims on others. Instead, he would simply maintain that assertions of orthodoxy were always epistemically unjustified. It is plausible to interpret the claim "saying just nothing at all" as an argument that an assertion of orthodoxy *does not speak to* the question of whether John may impose his views on others. If Locke denies (i), he seems correct. The mere fact that John is orthodox does not entail that he has the authority to require that others obey his demands. However, Locke might be denying (ii). Perhaps if John is justified in believing that he is orthodox he can permissibly impose his theological views on others. But the move from (i) to (ii) changes little. Neither a claim to be correct nor a claim to be epistemically justified is sufficient to show that John has authority over those to whom he addresses his demands.

Locke should not be interpreted as endorsing (iii). At other points in the *Letter*, Locke holds that if some religious point or other could be made manifest to the common reason of all, then it would then be an appropriate object of policy. For instance, in a doctrinal dispute between two churches, there is no "judge ... by whose sentence [the truth of doctrine] can be determined. The decision of that question belongs only to the Supreme Judge of all men, to whom also alone belongs the punishment of the erroneous."⁸ Accordingly, if a judge cannot determine which party to a doctrinal dispute is correct (such as God), the winner cannot have the right to rule. It is because disputants cannot agree as to whether John is orthodox that he has no authority over others. For

⁸ Locke 2003, p. 225.

Locke, if a judge could rationally determine John's correctness in accord with some set of impartial, publicly-available standards, then he may have political authority—but this is precisely what Locke denies we have in the case of religion. This interpretation fits with Locke's understanding of right judgment elsewhere in his work.⁹ In the state of nature, individuals disagree about who has the correct interpretation of the natural law. Locke infers from the negative consequences of these disagreements, that rational and reasonable citizens will appeal their decision to an impartial judge. Following Locke in the *Second Treatise*, John can only acquire political authority via a judgment by a body authorizes to speak as the voice of the reasoning of the public as a whole. If John and Reba disagree about John's orthodoxy, John can only acquire orthodoxy if decided by a judge whose legitimacy Reba recognizes. To put it another way, John's orthodoxy can only obligate Reba to obey him if his orthodoxy can be justified *to Reba*. To have bona fide authority, John must first appeal to common rational standards recognized by all. If all can see the rationale for a particular policy, or at least can reasonably be expected to see it, then the policy is the proper subject of legislation. The problem with doctrinal disputes, on Locke's view, is that they cannot be resolved by common reason. Thus doctrinal claims cannot ground political authority not because they are false or because no one can know that they are true, but because disagreement about theological doctrine cannot be resolved in a way that all can recognize. Locke does not deny that one can know to a high degree of probability whether certain theological doctrines are true.¹⁰ But this high degree of personal justification is irrelevant to determining whether John has authority over Reba. The fact that John is correct or in possession of a high degree of epistemic justification does not imply that John's view can be justified in terms of commonly recognized reasons. What is epistemically justified to John is not necessarily what is epistemically justified to all.¹¹

The point of this bit of Locke exegesis is to draw out an intuition, namely that valid claims of authority must be addressed to those on whom the claims are made. The force of "every church is orthodox to itself" is precisely that when John claims that he is right, he fails to address the reasoning of all. Only God can determine who is right—theological disputes between John and those he would control will prove intractable. However, even if theological disputes were tractable,

⁹ The passage I refer to is in the *Second Treatise*, where Locke discusses the social challenges posed by the problem of private judgment. See *ibid.*, p. 109.

¹⁰ Locke goes to great lengths to show that he knows some religious claims are true in *The Reasonableness of Christianity*. See Locke and Ramsey 1958. However, when it comes to revealed religion, we can only have varying degrees of probable belief. Locke and Nidditch 1975, pp. 688-696.

¹¹ Locke later states that even "if it could be manifest which of these two dissenting churches [was] in the right, there would not accrue thereby unto the orthodox any right of destroying the other." But in this case, Locke is defending the "two kingdoms" or "two rules" doctrine that God has vested coercive power in the state alone. Locke 2003, p. 226.

Locke's point would remain: there is no political authority apart from address to the reason of others. John cannot obligate Reba unless he can justify his claim to Reba by her own lights. The notion that authority requires interpersonal address lies at the heart of liberalism. Political legitimacy and authority requires reasoning from the standpoint of others. *This* is the Liberal Idea.

I.iii Equality, Respect and Reasonable Pluralism

The Liberal Idea implies particular conceptions of two of political philosophy's core concepts—*equality* and *respect*. All modern political theories assume that human beings are somehow equal and that all deserve respect. The social contract tradition and its descendant, the public reason liberal tradition, share an interpretation of these two core political concepts. For Hobbes and Locke, for instance, human beings are equal in that they are naturally in a state of natural liberty vis-à-vis one another, where they are entitled to act on their private judgment. As Locke claims, "The *natural liberty* of man is to be free from any superior power on earth, and not to be under the will or legislative authority of man, but to have only the law of nature for his rule."¹² Persons are equals in the sense that no one is born the natural authority over another. No one gets to be the boss for free. Instead, the liberal tradition that follows Hobbes and Locke holds that genuine authority must be justified from the standpoint of others. In so holding, liberalism presumes an interpretation of the idea of human equality. Rawls often claims that the liberal tradition assumes that persons are not merely naturally equal but "free and equal."¹³ Freedom and equality are derived from the same root. Just as individuals are equal in that no one has natural authority over others, they are free because they have a natural liberty to act in accord with their private judgments. These ideas of natural freedom and equality are frequently grounded in the Kantian ideal of *respect for persons*. On the Kantian versions of the social contract, especially public reason liberal interpretations, treating others with respect requires treating them as free and equal by justifying our demands on them on their own terms. We treat others with respect when we acknowledge their natural freedom and equality by acknowledging them as subjects of justification. The demands we make of our equals and the coercive actions we take against them require acknowledging them as those to whom a justification is owed.

¹² Ibid., p. 110.

¹³ Rawls 1971, pp. 225-226.

The social contract tradition is often read as recognizing what John Rawls calls “the fact of reasonable pluralism.”¹⁴ Reasonable pluralism is the state of a social world that obtains when rational, honest and thoughtful individuals disagree about even the most significant matters in life. Rawls believed that reasonable pluralism was the natural “outcome of the free exercise of free human reason under conditions of liberty.”¹⁵ Rawls argued that the cause of reasonable pluralism were the “burdens of judgment” or the sources of reasonable disagreement between persons that seem to be ineliminable.¹⁶ Among these burdens are the naturally conflicting and complex evidence we have for our differing views, the challenges faced when weighing differing considerations to arrive at common judgments, the naturally vague and indeterminate political concepts we must employ in discussion.¹⁷

The fact of reasonable pluralism motivates our need to justify ourselves to others by their own lights. The requirement of interpersonal justification would be obscured without reasonable pluralism, for reasonable and rational individuals would always agree and so no justifications for coercion would not need to be offered. But since free societies all find themselves in a state of reasonable pluralism, we must face the fact that many of our fellow citizens will substantially disagree with us about our conceptions of the good life, of justice and so on. The role of reasonable pluralism within public reason liberalism is therefore to reveal and motivate the need for addressing claims of authority to others.

I.iv The Standpoint of Others

The ideal of reasoning from the standpoint of others plays a pivotal role within liberal political theory in the late 20th and early 21st centuries. As we have seen, reasoning from the standpoint of others shifts political justification from a concern with truth to a concern with a form of internal justification to all. It is for this reason that Rawls claimed that a conception of justice appropriate to citizens of liberal democratic societies is political, not metaphysical.¹⁸ For the public reason tradition, political justification does not appeal directly to truth but strives to be compatible with common human reason despite disagreement about what is true. For centuries, the social contract tradition

¹⁴ Rawls 2005, p. xviii.

¹⁵ Ibid., p. 144.

¹⁶ Ibid., p. 54.

¹⁷ Ibid., p. 56-57.

¹⁸ Ibid., p. 10.

has concerned itself with justifying political authority without appealing to a comprehensive theology. This is quite clear in Hobbes, Locke, Rousseau and Kant. Public reason liberals extend this concern to justify political authority independent of comprehensive moral and political doctrines that aspires to truth. It is for this reason that Rawls describes his political liberalism as an extension of “the principle of toleration to philosophy itself.”¹⁹ For liberals like Locke and Rawls, political justification requires that we take account of what others believe they have reason to do. Political authority requires a standard of *public* judgment or a method of justifying authority in politics that takes account of what all believe they have reason to do. The distinction between private and public judgment is the distinction between judging a coercive law or policy to be authoritative based on one’s own epistemic standards versus common epistemic standards, the reasoning of citizens as a whole.

I will refer to the practice of considering the reasons of others in order to achieve political justification in terms of *perspective taking*. In our personal lives, we take a *first-personal perspective*. We consider our personal reasons are and act on them. But when we consider political justification, our perspective must extend to include the perspective of others. Thus we must take a *second-person perspective*, where we consider not only our own reasons but the reasons of others.²⁰ Throughout the dissertation, I will develop the contrast between the first and second-personal standpoints in politics and social morality generally. I shall argue throughout Part I of the dissertation that the contrast between the first and the second-person standpoints makes sense of the debate over the proper role of religious conviction in politics. Public reason liberals emphasize that political justification requires taking a second-person point of view. The reason to take the standpoint of others is the realization—following Locke—that even if our private judgments about what morality requires are true and personally justified, they cannot render authoritative our claims on others.

II: The Presumption in Favor of Liberty and Public Reason

II.i The Presumption in Favor of Liberty

Since much of this dissertation focuses on the idea of political and moral authority, we must say what authority consists in. I take it that John has authority over Reba just when John has a *claim-right*

¹⁹ Ibid.

²⁰ The term “second-person perspective” is borrowed from Darwall 2006. I analyze Darwall’s idea in Ch.3.

that Reba performs certain actions under the relevant conditions. On the traditional Hohfeldian scheme, A has a claim that B φ if and only if B has a duty to A to φ .²¹ But by definition, the existence of a claim-right entails the existence of a *liberty-right*. A has a liberty-right or a privilege to φ if and only if A has no duty not to φ . So consider: If John has a claim-right that Reba mow his lawn, then Reba has a liberty-right to mow John's lawn. Further, John must have a liberty-right to require Reba mow his lawn. Public reason liberals often mix claim-rights and liberty-rights when they discuss political justification. Sometimes public reason liberals are interested only in what they loosely refer to as legitimacy, which I take to mean that a governing body has legitimacy when it has the permission to coerce its subjects. Thus, if John governs Reba legitimately, he has a liberty-right to coerce Reba into taking certain actions even if Reba has no duty to obey. But whether we're concerned with authority or legitimacy, a public reason liberal must pursue a justification for John's liberty-right or permission to coerce. Consequently, I will confine myself to discussing how public reason liberals understand the justification of only one Hohfeldian incident—the permission to coerce. If we want authority, then we must justify John's claim-right to coerce, which goes beyond a liberty-right to coerce because John can require Reba to comply with his coercion. When I speak of the public justification of coercion throughout the dissertation I will mean *at least* the public justification of the liberty-right to coerce, though at other times I will focus on the claim-right to coerce. When I speak of social morality in Chapters 3 and 4, I will focus on the liberty-right and claim-right to interfere through the practice of significant social ostracism and blame.

Since this dissertation often speaks of the authority to coerce, we should also say something about what coercion is. That said, this dissertation does not require a unitary and well-defined conception of coercion to proceed, nor is there space to develop one. Some will worry that a political theory cannot rely on a controversial concept like coercion without defining it, but any political theory will rely in large part on some controversial concept. A careful conceptual analysis of coercion is required for a full account of the Liberal Idea but I prefer to rely on an intuitive notion of coercion for our purposes. This notion of coercion includes core cases, those cases of coercion that any adequate conceptual analysis of coercion must account for. These cases include deliberately causing bodily harm, using physical force or the threat of force to regulate the actions and speech of individuals, imposing constitutional restrictions on a society, many forms of legislation, and so on. I will assume that coercion takes the form of actions and threats to engage in such actions. That actions can coerce is easy to see. If you are busy building a house and I force you to hand over your

²¹ Wenar 2010.

building materials, I have coerced you. The same holds for threats in certain cases, such as when I threaten to engage in a coercive action unless you comply with a demand. Claims and demands exert a similar kind of pressure, such as when I make a claim on your time or demand that you help me. While it seems strong to say that claims coerce, they certainly interfere. This might seem strange, however. After all, if you make a demand of me, I can ignore you. The problem is that moral demands are ones that we already acknowledge as having force. Because we live in a common moral community, we mutually acknowledge the normative force of claims made on one another within that community. Claims on members of our moral community interfere because they threaten us with sanction. Thus our demands on others to follow the law constrain the reasons on which we may act. Claims on the behavior of others count as interference in part because they impose a risk that we will pay a social cost if we do not comply.²²

In liberal political theory, it is often assumed that there is a presumption against coercion or a *presumption in favor of liberty*.²³ That is, no one can claim to have a liberty-right to coerce others without a justification that changes someone's normative status. For instance, in Locke's case, coercing the religious choices of others must be justified to them in accordance with their own reasoning. Consequently, in the absence of such a justification, individuals have a natural liberty to resist, that is, a liberty-right to resist. The presumption against coercion claims that political morality contains an *asymmetry* between the permissibility of self-directed action and the permissibility of coercively interfering with the self-directed action of others.²⁴ Thus, self-directed action retains a kind of default position within political theory in that such action is permissible.²⁵ Persons have a liberty-right to act on their own reasons unless they are given some reason to the contrary.²⁶ The presumption against coercion also specifies appropriateness conditions for our coercive actions and claims. To draw an analogy, the presumption of innocence within trial law does not provide a reason to suppose that the defendant in a court case is innocent. Instead, it structures judicial practice. It specifies how a successful and justified trial must proceed. Unless the prosecutor can show that the defendant is guilty beyond a reasonable doubt, then she has failed to successfully prosecute a crime.

²² A "social cost" here is a threat of punishment. Not all costs are coercive, however. If John dies and this leads Reba to feel grief, he has imposed a cost on her, but he has not coerced her through his death or in dying.

²³ I take a presumption against coercion and a presumption in favor of liberty to be identical.

²⁴ By "political morality" I mean that part of morality concerned with the justification of coercion.

²⁵ By self-directed action, I mean something approaching Stanley Benn's conception of autarchy, which is quite a bit thinner than the standard liberal idea of autonomy. Being self-directed or moved by one's reasons is not the same as being autonomous. See Benn 1988, pp. 152-169.

²⁶ Until some coercive norm is justified, this liberty-right includes the right to coerce others, such that two individuals could have liberty-rights to coerce each other simultaneously.

The presumption against coercion, like the presumption in favor of innocence, structures the practice of the justification of coercion. When the presumption is met, coercion is justified and otherwise not. In this way, while the presumption against coercion is only one of many presumptions in political morality (such as a presumption in favor of promoting the general welfare), liberals give the presumption against coercion special status. The presumption against coercion is not a mere part of political morality, like other presumptions. Instead, the presumption against coercion must be met to generate political norms, or make them *bona fide* norms. In other words, only when a reason meets the presumption in favor of liberty can the interfering claim or coercive action that the reason supports be justified. For a putative political norm to be a norm at all, it must be justified through meeting the presumption.

Public reason liberalism contains a presumption in favor of *negative* liberty, the traditional conception of liberty as freedom from interference, being allowed to choose to act and believe as one sees fit. This conception of liberty is familiar, though some will resist understanding public reason liberalism in these terms. However, my claim is only that public reason liberal political theory contains a presumption in favor of negative liberty, *not* that other conceptions of liberty play no role in political justification. In fact, we will see below that positive liberty provides a method of meeting the presumption in favor of negative liberty. Consequently, public reason liberalism need not reject one concept of liberty for the other but can embrace them both. I shall further explain the moral basis of the presumption against coercion in Chapter 4.

The presumption in favor of liberty is arguably found in the works of various contemporary liberal political philosophers. Stanley Benn defends a presumption against interference in politics and morality, claiming that “the burden of justification falls on the interferer, not on the person interfered with.”²⁷ Joel Feinberg is another prominent defender of the “presumption in favor of liberty” which he defines as follows: “liberty should be the norm, coercion always needs some special justification.”²⁸ Further, John Rawls, the premier contemporary social contract theorist, appears to endorse a presumption in favor of negative liberty in his liberal principle of legitimacy: “Our exercise of political power is proper only when we sincerely believe that the reasons we offer for our political action may reasonably be accepted by other citizens as a justification of those actions.”²⁹

²⁷ Benn 1988, p. 87.

²⁸ Feinberg 1987, p. 9. Feinberg is talking about coercion here and Benn is discussing interference. I take the former to be a special case of the latter. For more, see Ch.4, I.

²⁹ Rawls 2005, p. xlvi.

Rawls does not use the phrase “coercion” in this formulation, but for Rawls, “political power is always coercive power.”³⁰ And on his view, the question of public justification for citizens is this: “When may citizens by their vote properly exercise their coercive political power over one another when fundamental questions are at stake?”³¹ Since political power is coercive power, we can understand Rawls as endorsing a presumption against the use of coercion. Here is another formulation of his principle of public justification:

Our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principle and ideals acceptable to their common human reason.³²

The conception of liberty at work in this principle, then, is the liberty we possess when we are not coerced, i.e. negative liberty. For Rawls, coercive political power is only legitimate when it is exercised in light of a publicly justified constitution; he does not require coercion to be justified at the level of an individual law. Instead, it must only be justified with respect to “constitutional essentials” or basic constitutional matters. But even if some other unit of coercion does is the object of public justification, coercive institutions must be. Thus a presumption against coercion is still part of the Rawlsian project. We can conclude, then, that the presumption in favor of liberty is a common commitment of contemporary liberals.³³

II.ii Public Reason and Public Justification

Public reason liberalism delineates how to meet the presumption in favor of liberty. The presumption is met when a *public justification* is given in the form of a *public reason*. Public reasons are “public” because they are taken to have force across a society or to some relevant subset of that society. Essential to the idea of a public reason is that public reasons can be recognized and endorsed by others.³⁴ They are reasons that address others in the Lockean sense described above: public reasons are those recognized as having force from multiple points of view. Not all public

³⁰ Ibid., p. 68. He says the same later: “Political power is always coercive power backed by the government’s use of sanctions ...” Ibid., p. 136. Also see, *ibid.*, p. 139, pp. 214-217.

³¹ Ibid., p. 217.

³² Ibid., p. 137.

³³ See Raz 1986, pp. 8-12 for a well-known criticism of this position.

³⁴ I will refine the idea of a public reason in Ch.4:II.ii.

reasons achieve public justification, however. Public reasons must override other considerations in order to achieve a public justification and thereby meet the presumption in favor of liberty. Rawls's conception of public reason follows the social contract tradition in conceiving of political authority and coercion as standing in need of justifications given from the standpoint of others. Public reason liberalism follows Rawls by showing that coercion can be endorsed from every reasonable point of view. The idea of public reason is thereby the culmination of a discussion about political life begun by Hobbes, Locke, Rousseau and Kant.

III: The Public Justification Principle

III.i The Liberty Principle

The foregoing discussion has reduced public reason liberalism to two fundamental principles which Gerald Gaus terms the Liberty Principle and the Public Justification Principle. The Liberty Principle is familiar:

The Liberty Principle: Liberty should be the norm ... coercion always needs some special justification.³⁵ Unjustified coercion is pro tanto wrong.

The claim that unjustified coercion is “pro tanto wrong” can be understood as a presumption against coercion. Coercion is presumed wrong in the absence of special justification. The presumption against coercion, the Liberty Principle, lies at the heart of the political authority.

III.ii The Public Justification Principle

The Liberty Principle specifies the conditions that must be satisfied for coercion to be justified. The Public Justification Principle specifies when these conditions are satisfied:

The Public Justification Principle: A coercive action C is justified if and only if each and every member of the public P has (a) conclusive reason(s) R to endorse C.

³⁵ Feinberg 1987, p. 9.

These formulations of the two principles hold throughout the dissertation. The Public Justification Principle is one of the dissertation's foundational ideas, so let us unpack the principle slowly. A coercive action is justified when it is permissible. Thus, as stated above, the Public Justification Principle, when satisfied, justifies at least a *liberty-right* to coerce. The Public Justification Principle, however, is required to justify our claims on one another and the coercion we use to enforce those claims. For this reason, I understand the Public Justification Principle as not only justifying John coercing Reba but as justifying Reba's duty to obey. Many philosophers will object to running these two ideas together.³⁶ However, public reason liberalism tries to explain how free and equal persons can have *authority* over one another. In other words, it tries to explain how equal persons must obey the demands their fellow citizens. What must be publicly justified then is a *claim-right* or a right by A to demand that B comply, entailing a liberty-right for A to coerce in turn.

The Public Justification Principle specifies conditions for justifying coercive *actions*. I take actions to be a broad class of events produced by individuals or societies that can occur once or repeatedly. These actions can constitute practices and these practices can comprise institutions. Thus, a law is an action because it denotes a practice of acting in a particular way. A regime is a set of actions required to maintain social institutions. The coercive action C can therefore be specified according to the subject of justification in question. Examples of coercive actions include single coercive acts, laws, policies, practices, the imposition of constitutional restraints, and the characteristic activities of entire regimes. These examples can all be understood as specifications of the object of justification, or the imposition that must be justified. They specify ranges of coercive acts, where, say, a coercive act is a single instance of coercion, and a constitution is a wide range of regular, coercive acts extended over a long period of time. Just as the public justification of a claim to coerce implies both a liberty-right and a duty, so does the public justification of a coercive action. If John is publicly justified in coercing Reba, John has a publicly justified permission to coerce Reba, and Reba has a publicly justified duty to comply with the coercive action in question.

III.iii Completing a Public Justification

The Public Justification Principle is the test for justifications and some proposed coercive actions fail the test. In this case, they are *defeated*.³⁷ But a coercive action can achieve public justification by

³⁶ For a discussion of the relationship between political authority and the duty to obey the law see, Raz 1986, pp. 99-105.

³⁷ For a discussion of defeaters, see Gaus 1996, p. 144. I discuss defeater reasons in detail in Ch.4, V.ii.

itself or as a member of a set of coercive actions. To illustrate, suppose that a society disagrees about which public health plan to adopt. Discussion reveals that a number of coercive proposals are regarded as justified by all and are thereby undefeated, such as a single-payer plan, a two-tiered plan and a basic safety net plan. Citizens regard all plans as superior to no plan. It is not initially clear how the Public Justification Principle handles such cases. Gerald Gaus distinguishes between cases where either one or more than one proposal is justified in terms of whether a proposal is *conclusively* justified. A conclusively justified proposal is one that defeats all other proposals, but in the case of the public health plans, all three proposals are *inconclusively* justified.³⁸ If all citizens of such a society regard one plan as superior to the others, then a polity should select the plan all prefer. If different citizens have different rankings, however, the problem becomes more complex. Citizens may find, for instance, that there is no one proposal that all regard as superior. In this case, public reason liberals appeal to *decision procedures* to select among inconclusively justified proposals or proposals that are members of an “eligible set”. If a decision procedure can be conclusively justified to all citizens then it can select among an eligible set of proposals and thereby generate a publicly justified social choice. The social contract theorists clearly gravitated toward justifying decision procedures rather than justifying each law individually. The same holds for Rawls, who sees public justification as applying only to constitutional essentials, which specify how social decisions are to be made. In this way, the social contract and public reason traditions recognize that disagreement even over justified proposals is so common that decision procedures must regularly be employed to resolve them. Examples include judge-made law, citizen votes and legislative votes. However, decision procedures can be conclusively and inconclusively justified themselves. Justified decision procedures must at least reliably select among an eligible set of inconclusively justified proposals.³⁹

III.iv The Unanimity Requirement

The Public Justification Principle specifies that “each and every” member of the public must have a conclusive reason to endorse some coercive action C. Some will find this unanimity requirement implausible. If just laws are only those that every single member of the public has conclusive reason to endorse, there may be no just laws. The worry arises from a concern about reasonable pluralism. Reasonable people, when they exercise reason freely, tend to disagree about even the most

³⁸ For discussions of inconclusiveness, see *ibid.*, p. 152 and Schwartzman 2004, p. 194.

³⁹ For a relevant discussion, see Gaus 2011, pp. 389-415.

fundamental matters in life.⁴⁰ If we require unanimity for public justification, we may find that for any proposed coercive law there is at least one person who does not have a conclusive reason.

Unanimity is problematic if members of the public are not *idealized* (IV.ii). However, under conditions of reasonable pluralism (conditions which members of liberal societies live under), individuals' reasons substantially diverge; accordingly, convergence will be rare. Yet public reason liberals maintain that idealization can vindicate some coercive laws because rational and reasonable individuals will converge on a public framework of coercion. Whether or not idealization produces convergence, public reason liberals must affirm the unanimity requirement. Without unanimous justification, the reasons of some members of the public will be ignored. Those citizens will thereby be excluded from full members in the moral community and such exclusion is disrespectful. Even if we cannot achieve unanimity in practice, an ideally just society still publicly justifies every coercive act to every reasonable individual. If public justification can be achieved, we should not settle for anything less. And if it cannot be achieved, then unanimous justification can still serve as an ideal.⁴¹

III.v Positive Freedom

We have already discussed the role of negative liberty within public reason liberalism: there is a presumption in favor of negative liberty that must be met by satisfying the Public Justification Principle. Arguably the public reason liberal tradition has a place for positive liberty as well. Positive liberty views vary somewhat, but all views somehow tie what it means to be free with what one has reason to do: "... those advocating positive liberty tie freedom very close to reason; a free person must be a person who acts according to reason rather than through impulse, superstition, or custom or out of ignorance."⁴² To be free is in some sense to be autonomous. Autonomy, however, is a complex concept and admits of many conceptions. I shall simply stipulate that I take autonomy to be understood as "the life of reason." An autonomous person does not have to be the author of every one of her rational commitments, ideals and the like. Instead, an autonomous person is one who acts on her *best reasons*. Susan Wolf calls this interpretation of the Kantian notion of moral autonomy the "Reason View".⁴³ Arguably Rousseau held a similar position, since on his view the

⁴⁰ We find this idea at the heart of Rawls's *Political Liberalism*. See Rawls 2005, p. 4.

⁴¹ Astute readers will wonder how citizens should be treated when we fall short of the ideal, but I cannot address this complex matter here.

⁴² Gaus 2000, pp. 95-96.

⁴³ Wolf 1990, p. 68.

point of the social contract is “To find a form of association that will defend and protect the person and goods of each associate with the full common force, and by means of which each, uniting with all, nevertheless *obey only himself and remain as free as before* [emphasis mine].”⁴⁴ A citizen can be coerced by her polity, and thereby lose negative freedom. But if she *self-legislates* the coercion, that is, if she imposes it only on herself, then she gains positive freedom because she acts on her best reasons.

The public reason liberal tradition does not clearly embrace a merely negative view of freedom or a merely positive view. Arguably Hobbes and Locke embraced a negative conception of liberty, whereas Rousseau and Kant embraced a positive conception of liberty (perhaps in addition to a negative conception). Rousseau and Kant would maintain that when a public justification is achieved, citizens remain “as free as before” despite being coerced. Thus, when the presumption in favor of negative liberty is met, public reason liberals who follow Kant and Rousseau will claim that those coerced retain their freedom since they are acting in accord with Reason. I will follow Rousseau and Kant in claiming that when the presumption in favor of liberty is met by a conclusive public justification, an individual coerced in accord with a publicly justified proposal or norm remains free in the positive sense even if her negative liberty is restricted.

IV: Reasons and Idealization

IV.i Reasons

In this dissertation, the ideas of reasons and the proper specification of the public (variables R and P in the Public Justification Principle) take center stage. I will argue that a specific value for these two variables can resolve the conflict between public reason liberals and integralists. Part II analyzes these two elements in great detail, but some preliminary comments are in order.

Public reason liberals typically hold that there are constraints on what count as justificatory reasons. Broadly speaking, there are two conceptions of what count as permissible reasons, R: the *consensus* conception of public justification and the *convergence* conception. Fred D’Agostino describes the distinction:

If both A and B share a reason R that makes a regime reasonable for them, then the justification of the regime is grounded in their *consensus* with respect to R. If A has a reason

⁴⁴ Rousseau 1997, pp. 49-50. Emphasis mine.

R_a that makes the regime reasonable for him, *and* B has a reason R_b that makes the regime reasonable for her, then the justification of the regime is based on *convergence* on it from separate points of view.⁴⁵

A convergence conception of reasons requires only that individuals have their individual reasons R to support a coercive action C , whereas the consensus conception requires that all permissible reasons R to support coercive action C share some crucial property, usually what I will call accessibility⁴⁶ or shareability.⁴⁷ The dominant conception of public reason understands consensus to require mutually shareable or accessible reasons.⁴⁸ Later I will refine this categorization (Pt.II, Ch.5); for now it is enough to point out that the consensus conception of reasons will be a key source of conflict between persons of faith and public reason liberals. Many public reason liberals defend what Christopher Eberle calls *doctrines of restraint*,⁴⁹ or norms that require citizens not to rely on their religious convictions in certain political contexts.⁵⁰ The motivations for such requirements are often that religious reasons are not thought to be accessible or shareable. In Chapter 6, I will argue that many religious reasons can meet both requirements.⁵¹ In Chapter 7, I defend a convergence conception of public justification which will remove the bone of contention between public reason liberals and their faith-friendly critics. We shall see throughout the dissertation that resolving the role of religion debate depends upon vindicating convergence over consensus. The key question is whether public reason requires justifying coercion to all based on some common set of reasons or whether public reason liberalism is compatible with justification to each person on their own terms, even if they share no reasons.

IV.ii Members of the Public and Idealization

Variable P in the Public Justification Principle refers to the subjects of the justification of a particular coercive action. Members of the public are all those to whom a public justification is owed.⁵² When a

⁴⁵ D'Agostino 1996, p. 30.

⁴⁶ Christopher Eberle has already outlined them in one place. Eberle 2002, pp. 252-286.

⁴⁷ Some claim reasons are essentially shareable, see Nagel 1986, p. 178. Scanlon 1998, p. 178. Korsgaard 1996, p. 135.

⁴⁸ For defenses of convergence, see Gaus and Vallier 2009, pp. 35, 56 and D'Agostino 1996, p. 30.

⁴⁹ Eberle 2002, p. 109.

⁵⁰ For a few clear examples, see Greene 1994, p. 659, Fish 1996, p. 22.

⁵¹ Eberle makes a similar argument. See Eberle 2002, pp. 234-293.

⁵² Typically, a justification is owed to anyone that will be coerced, but in other circumstances the subjects of justification may be larger than the set of people coerced. For example, some laws may be publicly justified to all that bar coercion of

justification for a federal law is offered, it must be offered to all those to all citizens represented by the federal government.⁵³ But once the outer bounds of the relevant members of the public are set, public reason liberals specify how reasons are to be ascribed to the relevant public. Specifically, public reason liberals often engage in *idealization* in order to determine how to understand how the beliefs and values of citizens enter into the process of public justification. Idealization is the process of revising citizens' beliefs and desires to generate a conception of citizens' reasons based on good information and consistent beliefs and values. Political philosophers idealize because it seems plausible to hold that an individual's true reasons are not based on poor information, poor reasoning or incoherent beliefs and desires.

Idealization covers deliberative conditions for social decision-making among different citizens, the conditions under which members of the public decide whether a proposal is justified. For Rawls, the veil of ignorance models the reasonableness of the parties to the original position by holding back information from their deliberation that would impugn their impartiality.⁵⁴ Jürgen Habermas prefers deliberative conditions without such abstract idealization, instead favoring constraining the form of discourse between actual individuals.⁵⁵ David Gauthier prefers a bargaining scenario where individuals know the characteristics of persons but are unaware of which person they are.⁵⁶ Under most versions of public reason liberalism, members of the public deliberate over large increments of coercion, such as constitutional level decisions. But for deliberation over small increments of coercion, public reason liberals rarely work with abstract deliberative conditions. Once principles of justice and basic institutions are justified, the justification of smaller increments of coercion is often removed from the deliberative situation and left to actual persons.

Idealization must be understood in terms of citizens' *belief-value sets*. A citizen's belief-value set is the set of all of her beliefs, desires, goals, and plans, i.e., everything she thinks and wants.⁵⁷ A citizen is idealized when her belief-value set is altered by changing one or more of her beliefs, desires, goals and/or plans according to some criteria. Theorists may give her relevant information or require that she be fully rational, both of which will alter her belief-value set. To illustrate,

a particular group. If the group was to be coerced by a new law, the non-coerced may still have a claim against the law. I thank Chad Van Schoelandt for this point.

⁵³ We might worry that only those citizens who are coerced are subjects of justification but since the state is to represent the reasons and values of all citizens, membership in the public need not be so restricted.

⁵⁴ Rawls 2005, p. 104. Also see Rawls 1971, p. 118.

⁵⁵ Habermas 1999, p. 198. Deliberative conditions must be free, in particular, of inequalities of power.

⁵⁶ Gauthier 1986, p. 245.

⁵⁷ The idea of a belief-value set derives from Bernard Williams's conception of a subjective motivational set. See Williams 1981, p. 102.

consider a coercive policy for funding scientific research that will aid in reducing greenhouse gas emissions. Most reasonable but otherwise unidealized citizens will acknowledge the need to reduce greenhouse gases, but few will understand the connection between the scientific research and the reduction of greenhouse gases. Thus, their belief-value sets lack the information sufficient to develop beliefs about whether the scientific research is a worthwhile investment. As a result, their reason to reduce greenhouse gas emissions may not lead them to recognize a reason to fund the relevant scientific research. It is not unreasonable to assume, however, that citizens really do have a reason to fund the scientific research necessary to reduce greenhouse gas emissions if they are committed to reducing greenhouse gas emissions generally. As a result, public reason liberals have good reason to idealize citizens' belief-value sets to include the information necessary to make an informed decision about whether the research is worth the cost.

To motivate idealization, consider how the public reason project fares if it declines to idealize. Let us call this position “justificatory populism”, a view that requires public justification to proceed on the basis of reasons a person actually assents to.⁵⁸ Christopher Eberle holds that populist theories of public justification have two core features: “First, populist conceptions involve an *inclusive* construal of the public: the default populist position is that *each* citizen in a liberal democracy counts as a member of the public and therefore plays a role in determining whether some rationale counts as a public justification.”⁵⁹ Populist views also “take citizens as they are: the default populist position is that a rationale R counts as a public justification only if the members of the public find R acceptable in light of their existing [subjective motivational sets], irrespective of their epistemic pockmarks and doxastic defects.”⁶⁰ Populist conceptions of idealization are widely regarded as subject to enormous problems, since people can “withhold their assent because of obstinacy, selfishness, laziness, perversity, or confusion.”⁶¹ Public reason liberals therefore need a conception of idealization. The standard strategy of idealization is to move from reasons citizens accept to reasons that they *can* accept or that are *acceptable*. Public reason liberals argue that individuals have reasons when the reasons are ones they are rationally committed to under certain conditions. These “certain conditions” are idealization conditions. They include giving individuals more information, upgrading their reasoning capacities, and rendering the members of their belief-value sets more

⁵⁸ Gaus refers to this view as the “actual assent thesis.” See Gaus 1996, pp. 130-1. I shall treat the terms “actual assent thesis”, “justificatory populism” and “populism” as synonymous.

⁵⁹ Eberle 2002, p. 200.

⁶⁰ Ibid.

⁶¹ Gaus 1996, p. 121.

coherent.⁶² Idealization also often requires that individuals be *reasonable* or that their reasons be determined by what considerations these individuals would accept were they reasonable. Citizens are frequently taken to be reasonable in Rawls's sense, such that "they are prepared to offer one another fair terms of social cooperation" and recognize the burdens of judgment.⁶³ I take citizens to be reasonable when they will follow publicly justified moral rules so long as others do.⁶⁴ By idealizing, some public reason liberals believe themselves to be restricting membership in the public. But in my view justification is owed to all citizens where their reasons are understood in terms of the proper conception of idealization. Thus, no one is excluded from public justification which avoids running afoul of the unanimity requirement.

Frequently public reason liberals radically idealize in response to the weaknesses of justificatory populism, which involves altering belief-value sets so that they are fully rational, contain all relevant information and possess perfect coherence. Such idealization can often radically alter belief-value sets as a result and is criticized for doing so. Nicholas Wolterstorff, for instance, worries that what a citizen would believe were he radically idealized has nothing to do with "how he should actually be treated in the political sphere."⁶⁵ The more we separate idealized belief-value sets from the dingy, real-world belief-value sets of citizens, the less reason we have to count the idealization as generating a citizen's true reasons. At some point, the tie is severed and indeed, as Wolterstorff insists, the person being idealized has no reason to care about his counterpart's recommendations.

One goal of Part II is to articulate a theory of moderate idealization that avoids the pitfalls of radical idealization while simultaneously resisting ascribing reasons to citizens based on flawed reasoning. In my view, the role of religion debate can only be resolved with a theory of idealization that meets these conditions. Persons of faith, like Wolterstorff, have complained about Rawls's insistence that the parties to the original position not know their conception of the good.⁶⁶ The attempt by public liberals to "fiddle" with citizens' reasons has understandably met with skepticism by those who feel disrespected by the public reason liberal tradition. But in Chapter 10, I develop a conception of idealization I believe can satisfy these critics.

⁶² For discussion of idealization standards, see *ibid.*, pp. 30-42, Audi 2000, p. 67, Rawls 1971, p. 456 and Ch.8.

⁶³ Rawls 2005, p. xlii.

⁶⁴ For an analysis of reasonableness in partially game-theoretic terms, see Gaus 2008.

⁶⁵ Wolterstorff 2007, pp. 153-4.

⁶⁶ For a pointed version of the criticism, see Neal 1997, p. 697.

V: The Challenge Posed by the Liberal Idea

The Liberal Idea is that political justification under conditions of reasonable pluralism must be *public* justification. From Locke to Rawls, the social contract/public reason liberal tradition has maintained that a just society is one in which all state coercion is justified in terms that all persons can accept.⁶⁷ This liberal idea of public reason can be understood in Lockean terms: our claims to authority over others are not grounded in truth claims or claims that are merely justified to us. Being personally justified or being correct is not sufficient to obligate others; instead, to obligate another requires reasoning from her point of view. To treat each other as equals and with equal respect, we must take each other's perspectives by considering what impositions we all have reason to accept.

The surprising result of the Liberal Idea is that if we wish to achieve a justified system of political norms and rules, that is, a system of rules that obligate all persons, we must renounce our claim to achieve our values according to our own judgment alone. Instead, our choices, our projects, our pursuit of value must be restrained by considering the standpoint of others. Liberalism in this fashion sometimes requires that we not act according to our deepest convictions. Hence, it sometimes requires *disloyalty* to an individual's projects, plans and beliefs, including religious ones. This disloyalty raises the central challenge for the dissertation: If John cannot institutionalize his values, if he cannot act on his beliefs in public, then how can liberalism be fair to him? How can the requirement of public justification require John not to act according to his convictions? How can liberalism require violating individual integrity?

The next chapter converts these suggestions into arguments by examining the criticisms that faith-friendly philosophers advance against public reason liberal approaches to religion in politics. These criticisms have great force in part because liberalism is committed to recognizing them. The entire point of liberalism is to enable individuals with diverse aims to cooperate for mutual benefit and protect their individual liberty; how can restricting freedom with the requirement of public justification possibly increase or protect it? The religion and politics conflict is, in short, a problem intrinsic to liberalism because of its commitment to increase freedom by confining it. For the liberal, *this* is the religion and politics conflict, the tension between taking the standpoint of others and taking a first-person perspective on our reasons for action.

⁶⁷ Sometimes the Liberal Idea is restricted to state coercion, but it need not be. In Chs. 3 and 4 I will extend the Liberal Idea even into the moral realm.

Chapter 2: Objections of Faith

This dissertation aims to reconcile public reason liberals and their faith-friendly critics. Chapter 2 explains why reconciliation is required at all. The Liberal Idea requires that coercion be justified from the standpoint of others. Along with many, I view public reason liberalism as the most complete and compelling articulation of the Liberal Idea. And yet, public reason liberalism is said to fall victim to a surprisingly *illiberal* treatment of citizens of faith in liberal democratic societies. Specifically, it is often said that public reason liberalism requires citizens of faith to “privatize” their beliefs, that is, to repress their fundamental commitments when participating in politics. In doing so, public reason liberalism is thought to violate the integrity of citizens of faith. What I shall call the *integrity objection* is often derided but public reason liberals cannot dismiss it. Liberals have always aimed to produce a society where all are free; if some feel silenced or marginalized by liberal institutions and political norms, public reason liberals are committed to worry. Concerns about integrity threaten the foundations of public reason liberalism because they turn liberalism’s foundational commitment to individual liberty against its foundational commitment to reasoning from the standpoint of others.¹ Consequently, if public reason liberalism imposes significant costs on the integrity of citizens of faith, it is threatened with defeat on its own terms.

In this chapter, I will review the major objections leveled against public reason liberalism by its faith-friendly critics. I will show that the force of the objections depends upon the force of the integrity objection. We shall see that the integrity objection creates a tension between reasoning from the standpoint of others and living by one’s own lights, thereby pitting two of liberalism’s foundational commitments against one another. The tension between these two commitments will appear intractable. We will see precisely how the tension arises and why the religion and politics conflict does not resolve in favor of either side of the debate.

Let me stress again that while the dissertation focuses on the conflict between public reason and public faith, the problem it raises and the solution it offers extend to anyone who holds a particular sectarian doctrine, religious or not. That is, the integrity objection can be raised by secular citizens with strong ideological convictions, for instance. I follow Rawls in articulating the challenge to public reason as originating not from peculiarly religious views but from something like what

¹ Of course, integralist liberals advance many criticisms against public reason liberalism, though many do not depend on the nature of religious participation in civic life. That is, many of their criticisms are not distinctively religious criticisms.

Rawls calls “comprehensive doctrines”, religious or not. For Rawls, a comprehensive doctrine “includes conceptions of what is of value in human life, and ideals of personal character, as well as ideals of friendship and of familial and associational relationships, and much else that is to inform our conduct, and in the limit to our life as a whole,” doctrines not limited to religious views.² Rather than comprehensive doctrines, I focus on full or partial individual ideals, but the idea is similar (see Ch.4, IV). This dissertation focuses on religion as a *case* of a more general phenomenon—the tension between the integrity of many views and a publicly justified social morality. Chapter 2 isolates the practical features of integrity that raise challenges to public reason liberalism; while these features are to be found among religious commitments, they will also be found elsewhere.

Finally, I must emphasize the particular structure of the dialectic of the chapter. Chapter 2 provides an internal criticism of public reason liberalism by showing how integrity threatens public reason liberalism from the perspective of one who embraces the view. Its explicit aim is not to convince others that public reason liberalism is true. However, in revealing ways in which public reason liberals might avoid the objection, I suggest a method of strengthening the appeal of public reason liberalism to those who reject the view. In this way, I hope to speak to integralist liberals on behalf of public reason by taking their concerns seriously. The remainder of the dissertation is a testament to the power of their concerns, as I structure an alternative version of public reason so as to preserve integrity from assault. I believe that public reason will appear more attractive to its integralist critics as a result.

Chapter 2 has six sections. Section I classifies integralist objections to public reason liberalism and explains some key ideas and concepts needed to move the argument forward. Section II develops the integrity objection in detail. Section III analyzes the fairness objection to public reason liberalism and shows that its normative force depends upon the success of the integrity objections. Section IV assesses the denial of truth objection and, like Section III, shows that its normative force depends upon the integrity objection. Section V shows that the integrity objection has great force against the standard conception of public reason liberalism but that it is not unambiguously successful. Instead, we shall see that the power of integrity is somewhat mysterious. Section VI concludes with a suggestion for how the conflict between public reason liberals and their integralist critics can be resolved nonetheless.

² Rawls 2005, p. 13.

I: Preliminaries

I.i Unaddressed Objections

Many objections advanced by integralist liberals do not directly concern religious commitment (or its interesting practical features). I will sideline them for this reason. For instance, Nicholas Wolterstorff, a prominent integralist, has argued that public reason liberalism unfairly privileges the value of noncoercion over other competing values.³ One prominent objection that often appears to derive from the nature of religious commitment is the “incompleteness objection” to public reason liberalism, which holds that it lacks the resources to resolve many important political problems.⁴ I also decline to discuss concerns about liberal neutrality. Instead, I will simply interpret the ideal of liberal neutrality as the ideal of public justification. Publicly justified institutions in my view resemble the “neutral” institutions that liberals traditionally endorse.⁵ Another commonly discussed issue is whether religious reasons are appropriately “accessible” or “shareable”, i.e., whether they can be evaluated according to common evaluative standards or whether they can be shared by suitably idealized citizens respectively. I believe that concerns about requirements on reasons must be addressed in detail but I defer a discussion of these issues until Ch.6.

I.ii Definitions

Before classifying the objections, terms like “religion”, “religious commitment” and “faith” require working definitions. I take a “religion” to be a comprehensive doctrine that contains a core set of principles about the *supernatural* that proscribe practices, rituals, norms, beliefs and actions. A “religious commitment” is a commitment to affirm doctrines about the supernatural and to act in according with the proscribed practices of the religion in question. My conception of religious commitment is somewhat distinct from others in the literature. Michael Peery defines religious

³ Wolterstorff 2007, p. 139. This concern has been addressed indirectly in Ch.1:II when I showed that the presumption against coercion structures justificatory practices and is not a method of assigning a high weight to the value of noncoercion vis-à-vis other values. I will discuss this further in Ch.4:I.

⁴ Kent Greenawalt is best read as advancing this criticism. See Greenawalt 1988, pp. 98-172. For replies, see Reidy 2000 and Schwartzman 2004. My convergence view, defending in Ch.7, addresses the incompleteness objection by increasing the normative resources available to generate public justifications.

⁵ I follow Rawls in this: “I believe, however, that the term *neutrality* is unfortunate; some of its connotations are highly misleading, others suggest altogether impracticable principles. For this reason I have not used it before in this lectures.” Rawls 2005, p. 191.

commitments to include any answers about the ultimate meaningfulness of existence.⁶ Christopher Eberle employs the narrower notion of theistic content.⁷ My definition is narrower than Perry's and broader than Eberle's as it aims to include non-theistic religions while not treating secular humanism and various forms of nihilism as religious. Understanding religions as forms of supernaturalism captures the set of practices and belief systems typically designated as religious. It also allows me to apply the arguments of Chapter 2 to a wide range of positions.

Supernatural claims have two components and both deny naturalism. I understand naturalism as the conjunction of a metaphysical thesis and an epistemological thesis. Metaphysical naturalism holds that the set of extant entities is restricted to the set of entities describable by physics. Epistemological naturalism holds that the only propositions that can be known are those concerning the physical world and human psychology and its associated conceptual components, such as the laws of nature, mathematics and logic. Supernaturalism is a conjunction of two opposing theses, metaphysical supernaturalism and epistemological supernaturalism. Metaphysical supernaturalism holds that the set of entities that exist includes both the set describable by physics and a set of non-material, non-conceptual entities, such as immaterial beings like souls, angels, demons, gods or the God of classical theism. Epistemological supernaturalism holds that the propositions that can be known concern not merely the physical world and its associated conceptual components but also propositions concerning the beliefs, attitudes, will, thoughts, reasoning, commands, actions and/or testimony of at least one non-material, non-conceptual entity.

Faith must be carefully defined, even in light of the above definitions. Faith is frequently taken to essentially involve irrational beliefs or beliefs that lack evidence. Others hold that faith involves believing in propositions that go beyond reason that are not inconsistent with it. However, many religious believers hold that their beliefs about the supernatural do not go beyond reason or against it but rather are supported by it. To respect reasonable pluralism, public reason liberals cannot understand faith-based beliefs as essentially irrational or non-rational but must develop a conception of faith that represents the range of belief systems that claim to involve faith. Faith often concerns beliefs about the supernatural. For instance, one can have "faith in God" by believing that God exists. But faith frequently involves more than this. For instance, within the Christian tradition, merely affirming Christian doctrine is insufficient for salvation. Further, many non-religious persons have affirmed the existence of supernatural beings but they are not appropriately described as having

⁶ Perry 1991, p. 119.

⁷ Eberle 2002, p. 71.

faith. In the major monotheisms, faith is often compared to *trust*. In particular, it is trust in the promises that God makes to humanity. For instance, in Christianity, faith in Jesus is often understood as faith not only that He rose from the dead but that His promise to save those who believe is true. Faith also involves trusting that God's will is good. For this reason, I shall not give necessary and sufficient conditions for the idea of faith. But I will stipulate that a necessary condition for having faith is affirming doctrines about the supernatural and trusting in the testimony or revelation of supernatural beings. Faith might involve more than supernaturalist belief and trust but this conception will suffice for our purposes. We may now move to objections.

I.iii Classifying Objections

As far as I am aware, Patrick Neal has made the only attempt to systematically classify religious objections to public reason liberalism.⁸ Neal groups the objections into four categories: incompleteness objections, fairness objections, integrity objections and denial of truth objections. Neal's article focuses on criticisms of Rawls but his classification system generalizes. I have set aside the incompleteness objection but I reproduce Neal's description of the other three:

- (1) *The Fairness Objection* seeks to hoist the Rawlsian argument on its own petard. The claim is that rather than embodying the principle of fairness, the doctrine of public reason is itself an expression of unfairness insofar as it subjects religious citizens to restraints that are not applied to non-religious citizens.
- (2) *The Integrity Objection* is perhaps the most prevalent argument made against Rawls from the point of view of religious belief. The heart of this argument is that the Rawlsian idea asks the religious citizen to "split" himself in a way that does, or can do, damage to the moral and or religious integrity of the person. In being asked to conduct his political activity in accordance with public reason, and to treat his religious views as being fundamentally non-political, the citizen, so it may be claimed, is being asked to repress or deny a fundamental part of himself when he enters the public realm.
- (3) *The Denial of Truth Objection* may appear to be the same as the integrity argument, but it is different in important ways. This argument challenges public reason on the grounds that it

⁸ Neal 2009.

seems mistaken to require the citizen to avoid stating claims of truth as truth. The charge is advanced not so much in terms of alleged damage to the person (as in the integrity objection), but rather in terms of the social costs of encouraging hypocrisy and/or dissembling over the profession of truth as one sees it.⁹

Section II analyzes the integrity objection and offers an interpretation that differs from Neal. I address the fairness and denial of truth objections in Sections III and IV respectively. I argue that the latter two objections are closely related to the integrity objection and derive their normative force from it.

II: The Integrity Objection

II.i Integrity

The integrity objection depends on the philosophical concept of integrity. Despite being widely employed, the idea of integrity is somewhat obscure.¹⁰ Bernard Williams's conception of integrity is the most prominent. It is sometimes called the "identity" view.¹¹ Reba has identity integrity when she is true to her character, projects, plans and beliefs. Cheshire Calhoun claims that on this view, "integrity means fidelity to those projects and principles that are constitutive of one's core identity."¹² By a "principle" I mean a commitment to some comprehensive ideal of life or a corollary or implication of it.¹³ "Fidelity" to these ideals requires acting in concert with them and reliably not violating them.

Projects are more complex. Loren Lomasky describes a project as possessing three primary characteristics: (i) persistence, (ii) centrality and (iii) structure.¹⁴ Projects are persistent insofar as they extend over long periods of time. John's goal of going to McDonald's this afternoon is not a project as the goal does not persist long enough, but his goal of having a successful career in business possesses the requisite degree of persistence since it can extend over years and even decades. A

⁹ Ibid., p. 155-6.

¹⁰ Harry Frankfurt understands integrity as self-integration. See Frankfurt 1988. Others understand integrity as involving unconditional commitments. Someone has integrity when she will never violate her commitments. See Calhoun 1995. I will not discuss either of these views.

¹¹ Cox, La Caze and Levin 2008, Sec. 2.

¹² Calhoun 1995, p. 235.

¹³ I will develop a more comprehensive conception of integrity in Ch.4, IV.

¹⁴ Lomasky 1987, p. 26.

project has centrality when it “help[s] explain a life” or when it is the focus of many other goals.¹⁵ For instance, suppose that John’s major aim in life is to become a successful business owner. In this case, the project guides and orders his other actions and sub-plans. John might plan to expand his business by opening another store, a plan ordered by his broader aim of being a successful business owner. In this way, John’s project has structure because it guides and orders his other reasons and goals. We can contrast John with those that lack clear projects. A good literary example is Frederic Moreau from Gustave Flaubert’s *Sentimental Education*. Moreau is a romantic who is never satisfied with what he has. While in the beginning of the novel, he pursues Madame Arnoux with great passion, he periodically loses interest. His unwillingness to commit leaves him with four love interests, many of which overlap. His professional life displays the same pattern: he drops out of law school, quickly abandons his political aspirations and is unable to make firm, honest life choices. *Sentimental Education* shows Moreau as constantly busy yet lacking clear projects.¹⁶

Those with integrity are loyal to ideals that constitute their core identity. An individual ideal constitutes a person’s identity when it explains who she is. Martin Luther may rank among the most extreme examples of such integrity. Luther’s devotion to Christianity led him to take on the Roman Catholic Church, the most powerful political institutions of his time. When accused of heresy at the Diet of Wurms, Luther stood in front of the power elite of the Holy Roman Empire and famously refused to recant, saying “I cannot and will not retract anything, since it is neither safe nor right to go against conscience. May God help me. Amen.”¹⁷ Everyone believed that if Luther refused to recant his theological views, he would be convicted of heresy; only through the political machinations of his protector, Elector Frederick III, “Frederick the Wise,” did he escape. Luther’s behavior at the Diet of Wurms demonstrates that his belief in “justification by faith” constituted his core identity; it helps explain who Martin Luther was.

We should avoid adopting a particular conception of personal identity involved here, however. For instance, Roman Catholics may be concerned that the identity view of integrity is too individualistic and therefore cannot capture the concerns Catholics have about the role of religion in the public sphere. But the identity view can include corporate identities as well. It also bears emphasizing that those most loyal to their projects and principles do not necessarily have more integrity. An individual with integrity may have many small projects and care for her freedom to

¹⁵ Ibid.

¹⁶ Flaubert 2004.

¹⁷ Luther is often thought to have said, “Here I stand. I can do no other” at this meeting but scholars doubt this. See Brecht 1985, p. 460.

pursue them. Otherwise, extremely committed individuals may have an exceptionally dominant role in public reason liberal treatments of integrity-related issues.¹⁸ But what unites lives with integrity are that reasons issuing from the projects and principles affirmed by many citizens are particularly powerful and seem to therefore generate the strong claims to noninterference that liberalism has always sought to protect. Since integrity generates reasons for action in line with one's principles and projects, it appears that public reason liberals are committed to permitting those reasons to play a justificatory role (for more on this, see Ch.4, IV).

II.ii The Target of the Integrity Objection

Integralist liberals argue that public reason liberalism *per se* requires citizens of faith to violate their integrity. The integralist liberal then argues that the citizens of faith have no reason to violate their integrity and on these grounds rejects public reason liberalism. To successfully defeat public reason liberalism, then, integralists must show that public reason liberalism requires integrity violations. The standard strategy is to select some feature of public reason liberalism's conceptual structure and show how integrity violations are implied by it. To evaluate the implication, let us make it explicit.

Public reason liberalism is characterized by the twin theses of the Presumption in Favor of Liberty and the Public Justification Principle. The Public Justification Principle states:

The Public Justification Principle: A coercive action C is justified if and only if each and every member of the public P has (a) conclusive reason(s) R to endorse C.

Integralists like Nicholas Wolterstorff challenge standard interpretations of members of the public P and reasons R (Ch.1, IV). Integrity-based worries, however, typically conglomerate around the claim that justificatory reasons must be shareable or accessible (or meet a consensus standard). Explications of these standards vary.¹⁹ I define the standards as follows:

Accessibility: A's reason X is accessible iff members of the public are committed to regarding X as justified according to common evaluative standards.

¹⁸ Gaus 2003, p. 185.

¹⁹ For some examples of accessibility, see Greenawalt 1995, pp. 151-164. Brower 1994, p. 392. Nagel 1987, p. 232. For some examples of shareability, see Korsgaard 1996, p. 135 and Scanlon 1998, p. 73. I give details definitions of both standards in Ch.5:II.ii and Ch.5:III.i.

Accessibility Requirement: A's reason X can justify coercing members of the public only if X is accessible to them.

Shareability: A's reason X is shared iff X can be properly inferred from the belief-value sets of all members of the public.

Shareability Requirement: A's reason X can justify coercing members of the public only if X is shared among them.

I understand "common evaluative standards" to be standards of epistemic justification that are valid for all members of the public and that are recognized as such. They are standards of proper inference and evidence. In this way, I am appealing to Rawls's idea that public reason contains within itself the standards of inference and evidence appropriate for public justification, or what Rawls calls the "guidelines of inquiry".²⁰ Notice also that accessibility and shareability are themselves open to many interpretations. Determining whether a reason is accessible depends on a complex judgment about whether the reason is justified according to common evaluative standards. Likewise, whether a reason is shared depends on a complex judgment about which reasons can be inferred from the belief-value sets of all citizens.

Public reason liberals naturally assume that religious reasons are inaccessible and unshareable. A paradigm example of the assumption is advanced by Stanley Fish:

That is what neutrality means in the context of liberalism – a continual pushing away of orthodoxies, of beliefs not open to inquiry and correction – and that is why, in the name of neutrality, religious propositions must either be excluded from the marketplace or admitted only in ceremonial forms.²¹

Similarly with Amy Gutmann and Dennis Thompson:

But the primary reason why [appeals to Biblical authority in support of coercive laws] must be rejected as moral reasons [and thus as an appropriate basis for coercive laws] is that they

²⁰ Rawls 2005, p. 224.

²¹ Fish 1996, p. 22.

close off any possibility of publicly assessing or interpreting the content of the claims put forward by the authority.²²

Thus public reason liberals like Gutmann assume that accessibility and shareability requirements impose restraints on the behavior of citizens. Christopher Eberle terms the principles that specify the appropriate restrictions *doctrines of restraint*—norms that require citizens not act upon or offer certain reasons in certain political contexts.²³ It must be stressed that doctrines of restraint are strictly ethical principles, not norms which can be backed by coercion. They are instead akin to what Rawls calls a “duty of civility”.²⁴

Before analyzing some common doctrines of restraint, it should be noted that appeal to reasons can be restrained in two ways. First, members of the public may be required to limit the use of unshared or inaccessible bases of expressed political justification and action. Such principles govern the considerations upon which citizens may permissibly act or the reasons they may offer in public political conversation in attempting to justify coercion to others. In contrast, sometimes reasons are excluded from public justification as justifiers; they are barred from the set of reasons to which just institutions must be sensitive. *Principles of exclusion* regulate such reasons. To illustrate, compare a duty of civic virtue with the Establishment Clause of the US Constitution. A duty of civic virtue regulates how one should act or talk politically, whereas the Establishment Clause precludes certain considerations from counting as bases of coercion. Restraint concerns the *practice* of justification, exclusion the justification itself.

The distinction is important; a reason expressed in the public sphere need not impact the ratification or rejection of a proposal. The reasons John offers Reba to support a law may differ from the reasons Reba has to support the policy. To be consistent with the Public Justification Principle, laws must be sensitive to Reba’s reasons to support or reject a policy in order to be publicly justified; but, they need not be sensitive to the reasons John offers Reba in conversation or vice versa. Defending restraint requires showing that unrestrained political discourse and activity will result in the use of private reasons as bases of coercion. Since accessibility and shareability only require that citizens not be coerced on the basis of inaccessible or unshared reasons, they only apply to exclusion. Consequently, accessibility and shareability only imply restraint by means of exclusion.

²² Gutmann and Thompson 1996, p. 70.

²³ Eberle 2002, p. 68. The domain of restraint is often unclear; this paper takes restraint to apply to political conversation in electronic or print media, protest, activism and visible public conversation.

²⁴ Rawls 2005, p. 217.

To summarize, mainstream public reason liberals endorse consensus conceptions of reasons and take consensus to support principles of restraint. While both public reason liberals and their critics tightly associate public reason liberalism with restraint, moving from the former to the latter requires three substantive steps:

- (1) Public Reason Liberalism → Accessibility/Shareability Requirement
- (2) Accessibility/Shareability Requirement → Principles of Exclusion
- (3) Principles of Exclusion → Principles of Restraint

Step (1) requires a substantive argument that a consensus conception of public reasons is to be preferred to alternatives. Step (2) requires a substantive argument that many sectarian reasons (especially religious ones) are inaccessible or unshareable and so should be excluded from public justification. Defending step (3) requires bridging the gap between exclusion and restraint; once public reason liberals successfully defend a principle of exclusion they must explain how the principle of exclusion implies a civic or political ethic that mandates restraint. The integrity objection is advanced against public reason liberalism by means of principles of restraint; public reason liberalism is criticized only by implication. Supposing that all three implications hold, if the integralist can show that restraint violates integrity and that therefore citizens of faith have defeaters for doctrines of restraint, they will thereby acquire a defeater for public reason liberalism by *modus tollens*.²⁵ The target of the integrity objection is restraint.

II.iii Doctrines of Restraint

I will now discuss four of the most prominent doctrines of restraint.²⁶ First is Gerald Gaus's principle:

Gaus: A liberal citizen with deep religious convictions thus refrains from appealing to these convictions in public justifications ... because they do not provide other (i.e., secular) liberal citizens with reasons to accede to his demands.²⁷

²⁵ In Chapters 6 and 7, I will deny that the first two entailments hold. The third entailment raises a number of complex issues, but once the first two entailments have been rejected, the general line of argument in the dissertation will be sufficiently vindicated that we need not address these issues in detail.

²⁶ I will sometimes refer to these doctrines as “principles” of restraint for facility of prose.

On Gaus's principle of restraint, liberal yet religious citizens should not appeal to religious reasons because they cannot publicly justify the moral demands those religious reasons imply. Gaus's principle of restraint is rooted in his commitment to the public justification of morality as a whole. We may now turn to Charles Larmore:

Larmore: A commitment to treating others with equal respect forms the ultimate reason why in the face of disagreement we should keep the conversation going, and to do that, of course, we must retreat to neutral ground.²⁸

For Larmore, if we are to treat individuals with respect, we have to dialogue with them on the basis of shared considerations. Only in this way can citizens who disagree resolve their disagreement. Larmore's retreat to neutral ground implies that individuals should not appeal to considerations that their interlocutors cannot recognize as valid. This restriction applies to religious reasons on his view because these reasons are inherently sectarian. We may now turn to Robert Audi's principle:

Audi: One has a prima facie obligation not to advocate or support any law or public policy that restricts human conduct, unless one has, and is willing to offer, adequate secular reason for this advocacy or support [and] unless in advocating or supporting it one is sufficiently *motivated* by (normatively) adequate secular reason.²⁹

Audi's principle of restraint is the strongest of those I will examine. It requires that citizens not only refrain from offering religious reasons to citizens or dialogue with them on neutral grounds but also possess an adequate secular rationale and motivation for political action. Audi believes that his doctrine of restraint must be followed in order to respect the equality and liberty of others. Rawls offers our final principle of restraint:

Rawls: Reasonable comprehensive doctrines, religious or nonreligious, may be introduced in public political discussion at any time, provided that in due course proper political reasons –

²⁷ Gaus 1996, p. 142. Note that Gaus has modified his view, see Gaus and Vallier 2009.

²⁸ Larmore 1987, p. 67.

²⁹ Audi 2000, p. 86, 96.

and not reason given solely by comprehensive doctrines – are presented that are sufficient to support whatever the comprehensive doctrines introduced are said to support.³⁰

Rawls's principle is the weakest of the four. For Rawls, one may advance reasons deriving from citizens' comprehensive (including religious) doctrines in public political discussion. But the requirements of restraint can be satisfied *post-hoc*: the present use of comprehensive reasons is justified only if no public reason for a policy or law is forthcoming. Non-public reasons can be used in lieu of public reasons that are developing.

Principles of restraint vary in significance, from excluding religious bases of political action and dialogue to restraining the use of religious bases of political action from being the sole motivations for action for religious citizens. Nonetheless, restraint requires many citizens of faith to restrict otherwise natural applications of their convictions to politics. Principles of restraint are moral demands. For this reason, they interfere with those on whom the demand is made (Ch.1, II.i). Consequently, a *burden* is placed upon those to whom restraint applies. Specifically, restraint burdens citizens of faith by restricting the application of their core principles and projects to their political lives. The integrity objection is intended to rebut the case for the burden.

It must be said that principles of restraint are seldom restricted to religious reasons. Only Audi's principle distinguishes between religious and secular reasons.³¹ Gaus, Larmore and Rawls all advanced principles that apply to many non-religious positions and reasons. Thus restraint applies not merely to citizens of faith but to anyone deeply committed to a doctrine with unshareable or inaccessible values and concerns. This dissertation focuses on citizens of faith and thus contains no detailed discussion of citizens with non-religious integrity-based concerns, though the main line of argument here applies to integrity-related concerns they might raise. Further, focusing on religious objections helps to bring out the most public and acute form of conflict between public reason liberalism and integrity and can thereby bring out the essence of the entire role of religion debate in general.

II.iv The Integrity Objection—A Preliminary Statement

Consider a classic statement of the integrity objection from Nicholas Wolterstorff:

³⁰ Rawls 1997, pp. 764-5.

³¹ And now even Audi is moving away from this distinction. Audi 2009.

It belongs to the religious convictions of a good many religious people in our society that they ought to base their decisions concerning fundamental issues of justice on their religious convictions. They do not view it as an option whether or not to do so. It is their conviction that they ought to strive for wholeness, integrity, integration, in their lives ... etc. Their religion is not, for them, about something other than their social and political existence; it is also about their social and political existence.³²

Michael Perry similarly suggests that to “bracket” one’s religious convictions is tantamount to “[annihilating] essential aspects of one’s very self.”³³ The integrity objection maintains that the civil restraint or “privatization” required by public reason liberals is objectionable. Restraint is objectionable because it effectively bars citizens of faith from living integrated lives. Specifically, restraint is said to require citizens of faith to split themselves into public and private halves and only apply their religious convictions to their private lives. Yet many religious traditions not only inform one’s private conduct but provide their adherents with conceptions of justice that they are to use to shape their world. Many theistic citizens, for instance, consider themselves to be obligated by God to bring about justice in the world. From their perspective, then, the public reason liberal demands that they violate their duties to God. Thus, if such citizens of faith conceive of themselves in terms of their religious tradition, that is, they see themselves first and foremost as religious believers, then they will naturally apply their religious views to their lives as a whole. The public reason liberal, then, threatens many citizens of faith with something like (though perhaps less hyperbolic than) “[a]nnihilating essential aspects of one’s very self.”³⁴

Wolterstorff and Perry may appear to advance a psychological objection—that principles of restraint alienate religious persons from themselves by frustrating them, causing stress, etc. But the integrity objection is best interpreted as concerned with publicly justifying moral demands. The psychological interpretation faces difficulties, as public reason liberals could simply reply by telling their religious critics to toughen up. In fact, Stephen Macedo suggests that citizens of faith who wish to press their religious claims within public reason should instead “grow up!”³⁵ The psychological interpretation thereby makes the integrity objection smack of whining; the religious critic objects to restraint on the grounds that citizens of faith are psychologically fragile. While the fact that a norm

³² Wolterstorff 1997a, p. 105.

³³ Perry 1988, pp. 181-2.

³⁴ Ibid.

³⁵ Macedo 2000, p. 35.

imposes great psychological burdens is undoubtedly some grounds for criticism, integralists must explain why public reason liberals cannot appropriately reply as Macedo does. Why shouldn't citizens of faith simply alter their psychologies insofar as they can in order to conform to the canons of public reason?

If integralists reply that citizens of faith have a claim against integrity-violating demands because they have *no reason* to restrain themselves, then these critics have abandoned the psychological interpretation for a normative one. On this view, the integrity objection concerns the moral norms citizens of faith have reason to endorse, about which impositions they can *reasonably be required* to endure. Citizens may have defeater claims for principles of restraint even from the perspective of public reason. If restraint imposes the psychological burdens integralists reject, then citizens of faith will probably have reason to reject the moral demands that public reason liberals impose upon them. If so, the case for restraint is defeated *on the basis of the Public Justification Principle*. We can thereby see why public reason liberals should be at least somewhat concerned about restraint: because restraint is an imposition or burden, there is a presumption against it. If the presumption cannot be met, the public reason liberals must oppose restraint as unjustified. The point of liberal political norms is to enable citizens with diverse ideals to coexist. If liberal political norms make it impossible for citizens to embrace many different ideals to live with integrity, then those citizens have strong reason to reject public reason liberalism on its own terms. Note how similar public reason and integralist liberals are on this point. Both share a belief that, at some level, political morality contains a strong presumption that individuals be able to live out their lives by their own lights.³⁶ The power in the integralists' integrity objection is precisely that it turns the presumption in favor of liberty against its public reason liberal advocates.

A successful version of the integrity objection must do more than cite costs to integrity—impositions that bar citizens from acting on their integrity-related reasons—in order to undermine public reason liberalism. All political theories impose some integrity costs by requiring citizens not to act on at least some of their core reasons. The public reason tradition, for instance, has always averred that accepting some integrity costs will prove a net benefit to those who wish to preserve their integrity.³⁷ If citizens give up their claim to act on some of their integrity-related reasons, they will gain more liberty to act on other, perhaps more significant concerns. Wolterstorff and Perry suggest that the integrity costs imposed by public reason liberalism are *high* because they require

³⁶ This is not to say that integralists embrace the presumption in favor of liberty in precisely the same sense as described in Ch.1.

³⁷ Understand a “benefit” to include an improvement in the satisfaction of one’s obligations.

splitting the identities of persons of faith. To split an identity is to corner-off the social space in which individuals can act in accord with their own judgments. Thus, political theories that split identities prevent citizens from acting on their convictions in some vital domain of life. Since activity in that domain is (it is assumed) of great import, restraint threatens to alienate citizens from their values and principles.

Like integrity costs, split identities do not by themselves imply a problem for public reason liberalism. Liberal society presumably splits the identity of a mobster; mobsters characteristically use violence to get their way and liberal society demands that they not do so. A better example is the split between a person acting in her office as judge and her private concerns. Surely there is no problem with *this* public-private splintering. If a religious judge or juror claimed that the restrictions of her office violate her integrity because they require her to only deliberate on admissible evidence, many would rightly balk. Liberalism makes demands of citizens that, if followed, will inevitably make some forms of life impossible, such as being a mobster or being a judge who rules based on her private reasons; such restrictions are not a bad thing. The power in Wolterstorff and Perry's argument relies precisely on the disanalogy: the religious person seems burdened *without sufficient reason*.

Public reason liberals will undoubtedly reply that the reasons for restraint are powerful enough to justify imposing the integrity costs to which Wolterstorff and Perry object; the duty to treat others as free and equal, they will claim, trumps the integrity-related reasons of citizens of faith and requires privatization. Evaluating who is right may prove difficult. To strengthen their arguments, Wolterstorff and Perry might try to show *how* integrity costs are imposed when citizens abide by restraint in political practice and then explain how restraint threatens to split their identities. For this reason, let us turn to Paul Weithman's more developed and empirically oriented arguments to this effect.

II.v Paul Weithman's Version of the Integrity Objection

Paul Weithman argues that many American churches play a positive role in promoting democratic action and political participation. For instance, from the Civil Rights Era to the present day, black churches have played a fundamental role in introducing members of black communities to politics. Weithman claims that African-Americans are one among many cultural and ethnic groups whose members often develop their citizenship by participating in religious organizations. The good

achieved by such participation is “realized citizenship,” a full involvement in politics that is a “great social and political achievement.”³⁸ For Weithman, churches contribute to the “realization of citizenship, especially by the poor and minorities.” They serve as “venues of discussion,” educate citizens about policy and candidates, and teach them “organizational and parliamentary skills” applicable to politics. Citizens thereby achieve “a sense of self-worth” that is “correlated with political participation” which in turn helps them ally with their fellow citizens. Churches promote realized citizenship by providing “the ability to participate in and to recognize debate as *public* debate.”³⁹ Without this ability one cannot achieve realized citizenship and the well-being associated with it.

Weithman rejects principles of restraint in part because they close off viable avenues towards realized citizenship. Restraint thus threatens citizens’ social and political development and prevents them from learning to deliberate based on considerations that define their interests. While Weithman does not object to principles of restraint on explicitly integrity-based grounds, his complaint is related.⁴⁰ Churches aid the formation of their parishioners’ identities as citizens around their religious convictions. Weithman might claim that in such cases citizens of faith will have no reason to accept norms that require them to restrict the development of a religious identity on the grounds that obeying such norms will make it difficult to develop their identities as citizens. Again, many citizens receive their political education in church and think about political problems in religious terms; impeding citizens’ acting on religious considerations thereby discourages them from political participation. To illustrate, suppose that African-American churches during the Civil Rights Era had accepted principles of restraint and so refrained from publicly defending civil rights on religious grounds. Since many African-Americans attained realized-citizenship through public expressions of religious commitment mediated by their churches, principles of restraint would have likely limited these substantial achievements.

That said, no public reason liberal argues that citizens should not discuss their religious reasons or organize politically in church. Instead, they merely require that when voting or arguing in the public sphere citizens should rely primarily on non-religious considerations. For this reason, it is hard to see how identity-development and political participation could be substantially set back by restraint. Some burden may be placed on churches’ ability to promote political participation but it is

³⁸ Weithman 2002, p. 22.

³⁹ Ibid.

⁴⁰ It is for this reason that I will interpret Weithman as advancing an integrity objection, rather than complaining merely about the loss of the good of realized citizenship.

not clear how significant the burden would be in practice. Weithman is no doubt correct that principles of restraint *can* restrict identity-formation and political participation, but it is not clear that restraint *must* have this effect. Weithman's version of the integrity objection therefore seems to rely on complex sociological judgments based on inconclusive evidence. Accordingly, the strength of Weithman's integrity objection is unclear.

Nevertheless, if widely accepted, many principles of restraint would undoubtedly have some of the effects Weithman describes. If citizens recognize that acting on religious reasons in the public sphere is considered immoral or inappropriate, they will be less inclined to act in accord with their religious identities for fear of public sanction. Consequently, should they choose to participate in politics on religious terms, they must bear the costs of being alienated from some sectors of their society. Weithman's point can be vividly illustrated in a number of real-world cases. For the latter half of the 20th century, South Africa was ruled by an apartheid government whose brutal crimes killed millions and created many more refugees. In 1990 Nelson Mandela was freed from a several decade prison sentence for protesting apartheid; in 1994, he was elected President of South Africa and started to reverse the social and political damage caused by apartheid. In the process, he formed the South African Truth and Reconciliation Commission, led by Anglican Archbishop Desmond Tutu. The Truth and Reconciliation Commission (TRC) was designed to expose the human rights abuses of the apartheid regime; amnesty was offered to those who committed serious human rights abuses in exchange for public admission of their misdeeds. The TRC embraced a conception of *restorative* justice which realizes justice through reintegrating perpetrators into their communities through confession and forgiveness. While the commissioners were politically, ethnically and religiously diverse, the philosophy of the TRC had an explicitly Christian emphasis.

Throughout his book *No Future without Forgiveness*, Tutu stridently defends the restorative justice approach pursued by the TRC on Christian grounds, arguing that no human, no matter how wicked, should be given up on. Christian theology reminded Tutu that perpetrators "remained children of God with the capacity to repent, to be able to change."⁴¹ For Tutu, "It was theology that enabled me to assert that this *was* a moral universe. That theology undergirded my work in the TRC." When Tutu was challenged by journalists for bringing his Christian convictions into the commission, he remarked:

⁴¹ Tutu 1999, p. 87.

I told them I was a religious leader and had been chosen as who I was. I could not pretend I was someone else. I operated as who I was and that was accepted by the commission. It meant that theological and religious insights and perspectives would inform much of what we did and how we did it.⁴²

Let us reconsider Weithman's argument in light of Archbishop Tutu's example. Tutu defended the TRC on explicitly Christian terms in innumerable public places; his Christian commitments directly influenced his role as the leader of a prominent and public institution responsible for healing deep rifts in a nation. If the principles of restraint advocated by public reason liberals were widely acknowledged within South African society, Tutu's witness would have been substantially muted. Again, Tutu "could not pretend [he] was someone else."

Even Rawls's weaker principle threatens Tutu's accomplishments. Again, Rawls believe that when citizens are "discussing constitutional essentials and matters of basic justice we are not to appeal to comprehensive religious and philosophical doctrines."⁴³ Tutu could publicly reach other South Africans in terms of comprehensive Christian values, and sometimes only in those terms; thus in some cases, Tutu would not have been able to explain his political activities "in terms of a reasonable balance of political values."⁴⁴ Tutu brought many into the fight for social justice in South Africa who otherwise would have never found their political voice. Principles of restraint would have made this already difficult process more burdensome. Weithman argues that good Rawlsian citizens must be able to make "complex judgments ... to determine when public reasons are called for" and have the sophisticated ability to "respond appropriately when offered the right or wrong kinds of reasons."⁴⁵ These requirements are unduly onerous for the severely oppressed citizens for whom Tutu advocated. Public reason liberals should be impressed that those who endured so much had the capacity to engage in political dialogue on *any terms at all*. In South Africa, it seems clear that principles of restraint would have closed off many avenues towards realized citizenship, and thus to the development of associated political identities. The South African case exposes the considerable burdens imposed by asking citizens of faith to only advance reasons in line with political values and supplement their religious arguments with language that does not resonate with their convictions. Restraint might be less onerous for college-educated citizens of Western liberal democracies. But

⁴² Ibid., p. 82.

⁴³ Rawls 2005, p. 224-5.

⁴⁴ Ibid., p. 243.

⁴⁵ Weithman 2002, pp. 206-7.

when restraint is applied outside of this privileged group, its restrictiveness becomes rather obvious. And it seems obvious even outside of the particularly difficult circumstances of South Africans. We need merely assume that ordinary voters in liberal democracies face similar hurdles to the process of realizing citizenship, that is, hurdles involved in learning how to engage in the complex reasoning processes required by doctrines of restraint. Thus, the integrity objection seems to wield great force.

II.vi Christopher Eberle's Version of the Integrity Objection

Christopher Eberle also has a powerful version of the integrity objection that is worth examining.⁴⁶ Above (II.iv) I showed how the integrity objection can be understood as an internal criticism of public reason liberalism. That is, if public reason liberalism imposes significant integrity costs, public reason liberals should be concerned given their own foundational commitment to individual liberty. Eberle's integrity objection is internal in the same sense, though it is aimed specifically at the Rawlsian version of public reason liberalism.

Eberle's argument begins with a description of "a common sort of theist" whose religious commitments imply that she is subject to obligations to God "... irrespective of [her] feelings, desires, or thoughts *about* those obligations."⁴⁷ Theists must take their obligations to God as "their most important obligation, such that in case of conflict between their obligation and some other they must opt in favor of obedience to God. That is, they'll regard their obligation to God as *overriding*."⁴⁸ A theist must be prepared to do God's will come what may. Eberle claims that if public reason liberals are interested in treating a common sort of theist as free and equal, they cannot advocate restraint because restraint will place undue strains on theistic citizens.

To demonstrate his conclusion, Eberle employs Rawls's "strains of commitment" argument against the liberal principle of legitimacy. Rawls's liberal principle of legitimacy states that,

Our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason.⁴⁹

⁴⁶ If one worries about the non-ideal citizens employed in my interpretation of Weithman's version of the integrity objection, Eberle's version will apply to even many idealized citizens. I thank Chad Van Schoelandt for raising this point.

⁴⁷ Eberle 2002, p. 144.

⁴⁸ *Ibid.*, pp. 144-5.

⁴⁹ Rawls 2005, p. 137.

One immediate problem for Eberle's argument is its target. The liberal principle of legitimacy is not a principle of restraint but one of *exclusion*. Eberle wants to refute restraint, not exclusion. Rawls and Eberle believe that the liberal principle of legitimacy implies the principle of restraint discussed above (II.iii). Consequently, I will apply Eberle's argument to Rawls's doctrine of restraint rather than his principle of exclusion.

Eberle first argues that the common theist's reasons will be excluded from public justification by Rawls's original position and that this will in turn lead to an unacceptable degree of restraint. To show this, Eberle maintains that the Rawlsian parties to the original position must make two determinations. First, they must ascertain what the liberal principle of legitimacy "might require of them under foreseeable circumstances" and second, they must discern "whether theists will be able to abide by these requirements."⁵⁰ Parties to the original position will set limits on the political actions that citizens may take, and so the first determination is "easy": parties will be barred from imposing policies on those who have no reason to endorse them. The parties will also regard the theist as having reason to accept the principle. If some party to the original position asks himself what restrictions he would have to accept were he a theist, he will conclude that he cannot act on his religious convictions if he cannot "provide a public justification" for the coercion in question. Accordingly, the theist "must withhold his support" from such coercion.⁵¹

Yet if Eberle's party regards his potential obligations to God as overriding, as a common sort of theist would, he will reject obligations requiring him to disobey God. Consequently, it is "quite doubtful" that a common sort of theist will "be able to withhold his support from political commitments he takes to be mandated by God in cases when he lacks a public justification for those political commitments."⁵² Instead, theists should prefer Eberle's ideal of conscientious engagement to Rawls's principle of restraint. The ideal of conscientious engagement requires the *pursuit* of public justification, but should a sincere person of faith make an honest effort to publicly justify her proposal and fail, she may still engage in political advocacy based on her religious beliefs even if she believes that it will result in publicly unjustified coercion.⁵³ Eberle maintains that since some reasonable theists believe they have overriding duties to obey God, they must select political principles that permit political advocacy grounded in their religious convictions alone.

⁵⁰ Eberle 2002, pp. 143-4.

⁵¹ *Ibid.*, p. 144.

⁵² *Ibid.* I understand "support" here as referring to restraint.

⁵³ *Ibid.*, p. 104.

Mainstream public reason liberals will dismiss as unreasonable citizens who are unwilling to compromise.⁵⁴ In response, Eberle marshals Rawls's "strains of commitment" argument, an argument which may be understood as a kind of integrity objection.⁵⁵ For Rawls, the strains of commitment are "strains that arise in [a well-ordered] society between its requirements of justice and citizens' legitimate interests its just institutions allow"; one such strain obtains between "the political conception of justice and permissible comprehensive doctrines."⁵⁶ Eberle believes that the strains of commitment require Rawls to reject the imposition of burdens on reasonable theists. Political conceptions of justice like Rawls's justice as fairness may only demand that adherents of reasonable comprehensive doctrines sacrifice so much. Eberle concludes that if a political conception of justice and its associated principle of restraint require theists to violate their most sacred obligations, then by Rawls's own theory the political conception and its associated principle of restraint must be rejected. Eberle takes it that Rawls's political conception and his principle of restraint fall victim to precisely this problem.

Contra Eberle, Rawls anticipates Eberle's criticism when he recognizes that some religious sects "cannot acknowledge any principle at all for limiting their claims on one another" since "the duty to religious and divine law [is] absolute."⁵⁷ For Rawls, that a religious citizen thinks others should share her beliefs on pain of damnation is no justification for restricting the liberty of others. Those oppressed cannot be expected to "acquiesce in an inferior liberty" nor can "we ask them to recognize us as the proper interpreter of their religious duties or moral obligations."⁵⁸ Rawls asks not only how principles of justice can justify coercive actions but—going back to Chapter 1—how they *obligate* citizens to comply with such actions. Eberle does not directly address this issue. On Eberle's view, theists claim the right to coerce others based on their theistic obligations. But no party to the original position has a reason to accord others the right to coerce based on reasons to which he is not committed.

In reply, Eberle will undoubtedly claim theists have the right to coerce others based on their theistic commitments because *God* has such a right and that He has delegated it to theistic believers.

⁵⁴ For Rawls, being reasonable essentially involves "a readiness to meet others halfway." Rawls 2005, p. 163. On this view, it is unclear whether Eberle's common sort of theist has this disposition to the appropriate degree. This is not to say they cannot.

⁵⁵ There is some dispute as to whether the Rawlsian concerns about the strains of commitment are normative or merely claims about what individuals can psychologically endure. Eberle takes the strains of commitment argument as a normative concern, so I will assess his argument in these terms. Rawls 1971, p. 153.

⁵⁶ Rawls 2005, p. 17.

⁵⁷ Rawls 1971, p. 182.

⁵⁸ *Ibid.*, pp. 182-3.

But theists in liberal societies rarely dismiss acknowledged moral requirements when they think God requires otherwise. Instead, they normally conclude that God would want them to follow their best understanding of political morality. So if the theist believes that political morality requires restraint, she will conclude that God wills restraint in the absence of countervailing evidence. Thus the theist will not only have reason to think that God wants her to obey social morality but that He has given her the capacity to determine which norms are justified and which are not. If theists determine God's will based upon what they consider to be moral, and if public reason liberals are right about how coercion can be justified, then reasonable theists will evaluate whether restraint is justified to them on the same grounds as anyone else. To be sure, they will maintain that public reason liberals can ask too much of them, but this point is not contested.

Eberle's integrity objection can be sharpened. For Rawls, religious reasons may only be included in public justification as "back-up" to reasons shared by those who endorse the political conception of justice. For instance, Rawls judges the religiously-based reasoning of the abolitionists and civil rights leaders permissible because "political" arguments against slavery and discrimination were forthcoming. The religious reasons they advanced may have been necessary to "give sufficient strength to the political conception to be subsequently realized."⁵⁹ Rawls remarks that the abolitionists could defend their religious language on the grounds that it "supported political values of freedom and equality for all, but that given the comprehensive ... doctrines ... in their day, it was necessary to invoke the comprehensive grounds on which those values were widely seen to rest." In defense of Martin Luther King, Jr.'s religious language, Rawls claims that King also spoke according to the principles of the US Constitution.⁶⁰ For Rawls, "religious doctrines clearly underlie King's views and are important in his appeals. Yet they are expressed in general terms."⁶¹

Notice that Rawls is trying to *excuse* King's language. For Rawls, there is something *prima facie* suspect about the religious language King and the abolitionists used to advance their political views. If Martin Luther King had not spoken in "general terms", then citizens would have good grounds for condemning King's defense of racial equality. Rawls only declares the abolitionists innocent because their language was necessary given the comprehensive values of their day. Religious language must be expressed in terms of the political conception of justice. Comprehensive reasons may only be used by individuals and groups attempting to integrate the political conception into their comprehensive doctrines. Thus, a theist has a civic duty to appeal to political values apart

⁵⁹ Rawls 2005, p. 251.

⁶⁰ *Ibid.*, p. 250.

⁶¹ *Ibid.*

from theistic values. It is not enough to speak in theistic language; any political activism he engages in must be on political terms. For Rawls, if the theist has religious reasons that *oppose* the political conception of justice, then acting on those reasons or advanced them in politics would be a mark of bad citizenship.

Eberle's integrity objection has great force against these views. Asking citizens of faith to only advance religious reasons in line with political values and supplement their arguments with language that does not resonate with their convictions is a substantial burden. People who come to political activism through their religious beliefs will have to second-guess themselves and develop the ability to decide whether acting on their convictions is appropriate based on whether their motivations and reasons are sufficiently public. If so, parties to the original position will have good reason to reject restraint. They will judge that restraint would unjustifiably burden them were they theists.

III: The Fairness Objection

The integrity objection is powerful. We will examine just how powerful it is below (V). But we may now turn to the second integralist objection raised against public reason liberalism, the *fairness objection*. The fairness objection holds that public reason liberalism imposes greater burdens on citizens of faith than secular citizens. Therefore, fairness objectors conclude, public reason liberalism treats citizens of faith unequally. The fairness objection is in some ways the articulation of the sense that many defenders of citizens of faith have that liberalism is biased against people of faith. Thus the fairness objection often proceeds as a deconstruction of public reason liberalism. It attempts to show that public reason liberalism's attempt to avoid imposing unjust demands on citizens fail and that it is no more impartial than any religious political ideology, only less honest about it.

III.i Michael Perry's Version of the Fairness Objection

Both Michael Perry and Nicholas Wolterstorff offer versions of the objection. Both versions proceed first by analyzing public reason liberalism's epistemic commitments and second by attempting to show that these commitments unjustifiably privilege secular persons. In criticizing Thomas Nagel and Bruce Ackerman's views on impartiality and neutrality within liberal political theory, Perry advances a standard account of the fairness objection:

[Nagel's theory] leaves me and many others at a serious disadvantage vis-à-vis Nagel and other devoutly secular intellectuals. In many political arguments Nagel might get to rely on all or most of his relevant beliefs, including his most important relevant beliefs, while I would get to rely on only some of my relevant beliefs, *not* including the most important ones: my religious convictions about human good.... therefore ... Nagel's practice of "impartial" political justification is not impartial, even though particular arguments yielded by the practice are impartial in the special sense of presupposing the authority of only widely accepted beliefs.⁶²

For Perry, Nagel's conception of public reason liberalism privileges secular beliefs despite advancing putatively impartial criteria for acceptable political justifications.⁶³ Perry is not arguing that Nagel's moral beliefs can enter into public justification merely because they are secular beliefs. The complaint is that Nagel's conception of impartiality is objectionable precisely because it results in an implausible double standard for religious and secular beliefs. Nagel might reply that the standards that result in privileging many secular beliefs are rationally justified standards and, consequently, the marginalization of religious beliefs is both justified and fair. Secular reasons are privileged by an impartial moral epistemology because they are *better* reasons. Perry replies by defending a coherentist conception of rationality. He holds that whether it is rational to affirm p is judged by the degree to which p fits with other beliefs one has. On his view, "the basic test ... for determining what beliefs it is rational for a person to accept is coherence with ... beliefs that are presently authoritative for him."⁶⁴ If coherentism is a plausible epistemological account for evaluating moral reasons and beliefs, then it looks as if the asymmetrical treatment of religious and secular beliefs cannot be sustained. Individuals' beliefs are rational to the extent that they cohere with other beliefs. Perry seems to argue that if so there is no reason to suppose that religious belief cannot be as rational as secular belief. The rationality of such beliefs depends on one's other beliefs. Perry contrasts his conception of political rationality with what Bernard Williams calls the "linear" model of rationality, which holds that the epistemological enterprise rests upon "the structure of knowledge on some favored class of statements."⁶⁵ Perry argues that public reason liberalism relies on the linear model,

⁶² Perry 1991, p. 15.

⁶³ Let "secular beliefs" denote reasons deriving from doctrines of a non-religious character.

⁶⁴ Perry 1991, p. 53.

⁶⁵ *Ibid.*, p. 54.

which privileges certain types of beliefs and reasoning, whereas on the true (coherentist) account of rationality, beliefs cannot be privileged in this way.⁶⁶

This objection seems problematic, though. Public reason liberalism does not depend upon rejecting a coherentist epistemology. In fact, coherentist approaches to political epistemology are common. Rawls is an exemplar coherentist in this regard, since his justificatory method of reflective equilibrium is coherentist in just the sense that Perry prefers.⁶⁷ In fact, Rawls argues that the political conception of justice (which he believes entails the liberal principle of legitimacy) is derived from a society's shared ideas. His original position abstraction is intended to render coherent the shared ideas of "you and me".⁶⁸ It is not unreasonable to think that the mainstream conception of public reason liberalism requires more of a citizen of faith than secular citizens, but Perry's is not the reason why.

III.ii Nicholas Wolterstorff's Version of the Fairness Objection

Like Perry, Wolterstorff holds that public reason liberalism's epistemic structure causes it to assign privileged epistemic status to secular beliefs. But Wolterstorff blames this inequality of status on what he calls its "independent source" epistemology.⁶⁹ According to Wolterstorff, on the public reason view, citizens "... are to base their political debate in the public space, and their political decisions, on the principle yielded by some source *independent of* any and all religious perspectives to be found in the society."⁷⁰ The independent source epistemology will therefore straightforwardly privilege secular beliefs over religious ones. Wolterstorff suggests that the case for this privilege is weak. But more importantly, the problem with the independent source view is that it unjustly restricts a religious citizen's free exercise of religion. Wolterstorff writes that,

To require of [citizens of faith] that they not base their decisions and discussions concerning political issues on their religion is to infringe, inequitably, on the free exercise of their religion. If they have to make a choice, they will make their decisions about constitutional

⁶⁶ That said, on the next page Perry denies that coherentism forbids privileging some beliefs. *Ibid.*, p. 55.

⁶⁷ Rawls 1971, pp. 18-19.

⁶⁸ Rawls 2005, p. 28.

⁶⁹ Audi and Wolterstorff 1997, p. 78.

⁷⁰ Wolterstorff 1997b, p. 166.

essentials and matters of basic justice on the basis of their religious convictions and make their decisions on more peripheral matters on other grounds⁷¹

The problem here is that the independent source view impinges on the liberty of citizens of faith in a way that it does not for secular citizens. Such unequal treatment is *prima facie* unfair and should cast the independent source view into doubt for that reason alone. But is public reason liberalism committed to an independent source epistemology?

Wolterstorff may initially appear to be on excellent ground in claiming that public reasons do not attain their positive epistemic status from any religious perspective, that is, public reasons derive from an independent source. Rawls claims that a political conception of justice must be “freestanding” in that its principles are determined on the basis of considerations not reliant on a particular comprehensive doctrine.⁷² That said, on the standard conception of public reason liberalism (that I reject), public justification must proceed according to reasons accessible or shareable with all. Such reasons are not thereby “independent” of religious perspectives. For Rawls, for instance, the political conception of justice is a “module” which must fit within each reasonable comprehensive doctrine.⁷³ Rawls’s view is that the political conception of justice must cohere with a reasonable citizen’s total set of beliefs; his view does not postulate an independence relation between religious beliefs and others. The first stage of Rawlsian political justification—the “pro tanto” stage—is freestanding, i.e., it is justified according to considered judgments shared by the perspective of “you and me”.⁷⁴ But the considered judgments used to construct the political conception are not therefore independent of comprehensive doctrines. Instead, the political conception is worked out from a subset of one’s commitments. This may smack of a kind of independence, but Rawls’s second stage of political justification belies independence entirely. For Rawls, “full justification” requires that the political conception of justice be justified from within the comprehensive doctrines of every reasonable person.⁷⁵ Thus, Rawlsian political justification does not justify political power independently of comprehensive doctrines, including religious ones.⁷⁶

Nonetheless, Wolterstorff’s concern about inequitable treatment still stands. Even if public reason liberalism is not rooted in an independent source view, it still imposes unequal burdens on

⁷¹ Wolterstorff 1997a, p. 105.

⁷² Rawls 2005, p. 10.

⁷³ *Ibid.*, p. 12.

⁷⁴ *Ibid.*, pp. 28.

⁷⁵ *Ibid.*, p. 386.

⁷⁶ For further discussion of Rawls’s three steps of political justification, see Ch.7, II.

citizens. Yet it is important to be clear about who is being imposed upon. Many citizens will find shared political ideas most salient and politically inspiring. However, other citizens will be most inspired and moved by ideas that are unique to their particular worldviews and traditions. Religious citizens are not necessarily more burdened than some secular citizens. That said, all public reason liberals acknowledge that secular considerations are more likely to be shared than religious considerations. And so, in effect, public reason liberalism seems to impose unequal burdens on citizens of faith, even if it does not do so in principle. Restraint thus seems unfair.

III.iii Assessing the Fairness Objection

Something seems right about the fairness objection but I argue that its normative force depends upon the success of the integrity objection. To show this, suppose that the burdens public reason liberals impose upon religious and secular citizens are indeed unequal. Public reason liberals can rectify the problem in two ways: (a) they can decrease the burdens imposed on religious citizens until the burdens are equal, or (b) they can increase the burdens imposed on secular citizens until the burdens are equal. It is fair to assume that Wolterstorff and Perry will prefer (a) to (b) because they prefer for citizens of faith to have more freedom rather than less. This preference is clear from the integrity objection. Now suppose that the fairness objection has normative force independent of the integrity. If so, it should generate *at least some reason* to prefer (b) to an inequality of burden. In other words, Wolterstorff and Perry will have some reason to reduce the freedom of secular persons despite securing no increase in the freedom of citizens of faith. Since Wolterstorff and Perry will likely reject this sort of leveling down, it appears then that inequality is not the only issue. Instead, it seems that the primary problem is that unjustified burdens are placed on citizens of faith. The power of the fairness objection seems to depend upon the integralist worry that public reason liberalism burdens the liberty of religious citizens by asking them to do what they have no reason to do.

IV: The Denial of Truth Objection

The denial of truth objection comes in stronger and weaker versions. The stronger version holds that public reason liberalism unjustifiably requires citizens to deny the truth of their comprehensive doctrines. The weaker version holds that public reason liberalism requires citizens to *treat* their

comprehensive doctrines as if they were false, or to act inconsistently with them. The objection focuses, in Neal's words:

... upon the apparent oddity of recommending that one not speak the truth as one sees it in certain political circumstances. Ordinarily, we count it a virtue of persons to express the truth of things as best they understand it. Insofar as public reason cuts against telling the truth, it confronts the potential critical force of this ordinary moral intuition that it is good to speak the truth.⁷⁷

In short, the denial of truth objection points to the oddity of public reason liberalism's insistence that citizens not base political action on the truth as they see it. In contrast, William Galston has argued that to treat others as free and equal, we are only obligated to offer others "what we take to be our true and best reasons for acting as we do."⁷⁸

The denial of truth objection typically targets Rawlsian political liberalism, a doctrine that explicitly commits itself to not appealing to the whole truth as citizens' see it when justifying political power to one another. Rawls faces up to this concern when he asks how it "can it be reasonable or rational, when basic matters are at stake, for citizens to appeal only to a public conception of justice and not to the whole truth as they see it?"⁷⁹ Rawls maintains that respect for persons requires that we appeal to reasonableness rather than truth when justifying coercion. But Rawls emphasizes that his theory does not require citizens to deny the truth of any reasonable comprehensive doctrine. Rawls's concern about truth claims arises when *the fact that the claim is true* is used to justify coercion or authority. Rawls argues that "holding a political conception as true, and for that reason alone the one suitable basis of public reason, is exclusive, even sectarian, and so likely to foster political division."⁸⁰ Rawls's point here is precisely the point of Chapter 1: truth claims by themselves cannot show that others are obligated to comply with our coercive actions or claims upon them. It thereby seems false to claim that political liberalism, and thus public reason liberalism generally, requires that citizens deny the truth as they see it or act inconsistently with it. Perhaps the denial of truth objection is that public reason liberalism requires citizens of faith to be disloyal to their principles, but this is a version of the integrity objection.

⁷⁷ Neal 2009, p. 153.

⁷⁸ Galston 1991, p. 109.

⁷⁹ Rawls 2005, p. 215.

⁸⁰ *Ibid.*, p. 129.

A critic may reply that principles of restraint and exclusion make truth claims about which propositions are relevant to public justification. She might then argue that the separation of reasons that justify coercion from those that do not requires her to deny her comprehensive doctrine in an objectionable fashion. It is true that public reason liberalism separates considerations that may enter public justification from those that do not. The question is whether principles of delineation themselves require citizens of faith to deny important truths. The answer depends on the faith in question and how restrictive public reason liberalism turns out to be. I argue in Part II that a proper understanding of reasons excludes very few reasons from entering into public justification despite the fact that a commitment to public justification will prevent many reasons from successfully publicly justifying coercion. Consequently, the version I develop should be immune to the denial of truth objection, though standard accounts of public reason liberalism may separate reasons in a more burdensome and thus objectionable fashion.

The question we must ask ourselves now is why the process of delineating between public and non-public reasons is objectionable. From what I can tell, it is objectionable because it imposes costs upon the integrity of citizens. The problem of forcing someone to deny the truth of her principles has nothing to do with whether those doctrines are true, as we would lodge the same complaint on behalf of someone with a false set of principles. Instead, the problem seems to be tied to integrity based concerns. Forcing someone to deny her principles is wrong, whether those principles are true or false. Why? One compelling explanation is that forcing her to do so requires her to violate her integrity.

V: The Mysterious Power of Integrity

The integralists seem correct: the mainstream version of public reason liberalism imposes significant restrictions on the integrity of citizens of faith by means of restraint. It may therefore appear—assuming the three entailments between the Public Justification Principle and restraint hold—that public reason liberalism is defeated. Public reason liberalism aims to treat citizens as free and equal under conditions of reasonable pluralism, thereby allowing citizens with different conceptions of the good the liberty to act on their projects and plans. Thus, a commitment to public justification seems to imply a commitment to respecting the integrity of citizens. If public reason liberalism requires unjustifiable restrictions on integrity, then its commitment to individual liberty seems to undermine itself.

Public reason liberals will protest, however. On the Rousseauian and Kantian versions of public reason liberalism, when moral demands are publicly justified, they are *self-legislated* (Ch.1, III.v). When we show that moral demands are justified to others by their own lights, then they are rationally committed to the coercion. In this case, morality no longer violates their liberty. For Rousseau, the social contract accomplishes this freedom-preserving trick:

To find a form of association that will defend and protect the person and goods of each associate with the full common force, and by means of which each, uniting with all, nevertheless *obeys only himself and remains as free as before* [emphasis mine].⁸¹

When coercion is publicly justified, it no longer restricts liberty. This is an ideal that lies at the heart of public reason liberalism.⁸² It is a traditional conception of positive freedom, where one is free or autonomous, when she “can decide for herself what to do and is not the slave of impulses, ignorance, error, past conditioning, or addiction. She is, literally, self-ruled.”⁸³

For the public reason liberal, if the consensus conception of reasons is the correct conception of public justification, then the norms entailed by consensus (like restraint) *cannot possibly* violate individual integrity. If John’s religious tradition informs his conception of justice and yet he self-legislates the privatization norm, then his best reasoning will render his religious commitments and commitment to privatization consistent. In this way, privatization will not limit John’s freedom because he imposes privatization on himself. For the public reason liberal, therefore, the integralist has made a fundamental conceptual error. Public reason liberalism simply specifies how to understand this condition of self-imposed freedom. There is a sense in which John’s personal projects and plans are restricted by principles of restraint but, again, they are principles which he imposes upon himself. John cannot complain that public reason liberals demand he keep his religion to himself when he himself self-legislates that demand.

The complaints of the mainstream public reason liberal to the contrary, the integralist is on to something. Restraint does restrict integrity and it seems to do so in an objectionable way.

⁸¹ Rousseau 1997, pp. 49-50.

⁸² While many public reason liberals do not explicitly embrace an ideal of positive freedom like Rousseau’s, they often speak in terms of achieving some form of autonomy, even if only the “full autonomy” of Rawls’s political liberalism. See Rawls 2005, pp. 77-80. On a Rousseauian or Kantian view to achieve autonomy just is to be free. So I think it is fair to say that some ideal of non-negative freedom like positive freedom or freedom as autonomy lies at the heart of even public reason liberalism.

⁸³ Gaus 2000, p. 83.

Undoubtedly integralists will not share the positive conception of liberty so characteristic of Rousseauian and Kantian conceptions of public reason liberalism. Perhaps then the division between public reason liberals and integralist liberals is not first and foremost a concern about religion in the public square but a disagreement about the *very idea of freedom*. And yet this seems wrong. Public reason liberals themselves are committed to respecting integrity and it still appears, the idea of positive liberty to the contrary, that they should be concerned about the integralists' complaint. The problem for the public reason liberal is that restraint appears to be heteronomous in the traditional sense. By insisting that citizens of faith not to appeal to the totality of their reasons, public reason liberals impose morality from the outside rather than showing that citizens of faith are committed to the norm by their own lights. The suggestion here is that complaints about restraint point to an internal criticism of public reason liberalism: *public reason liberals have the wrong understanding of public reason because the norms implied by their view seem heteronomous*. Their demands raise complaints of unfreedom and disintegration from sincere citizens of faith and their philosophical advocates. If public reason liberals listen to their critics, they should hear not mere whining but complaints of heteronomy. The integralist complaint is not best understood as a complaint that citizens of faith lack a mere negative liberty to speak as they please, but that they lack the positive liberty to live in concert with the totality of their reasons and commitments.⁸⁴

The irony in this is that public reason liberalism claims to not only preserve integrity through public justification but to promote it. Public reason liberals can be understood as making an offer to citizens: cede sufficient negative liberty to live in accord with publicly justified rules, and you will have more positive liberty as a result. Citizens have reason to accept the restrictions required to have civil society because society is a cooperative venture for mutual gain. By living together despite our differences, our lives will be vastly improved. In doing so we are able to extend our projects and plans. By accepting the restrictions involved in social life, we can have more integrity than we could under publicly unjustified regimes. Therefore the attraction of living an integrated life should draw us towards public reason liberalism, not push us away. Only public reason liberalism promises citizens a truly *moral* peace with others, a peace required for a just form of social stability where we can act upon and expand our projects and plans and live with fidelity to our principles.

So to conclude that integralist concerns refute or defeat public reason liberalism is premature. And this is precisely because public reason liberalism is already committed to respecting

⁸⁴ For the time being, I understand negative liberty broadly as the freedom one has when she is not interfered with in regards to some matter. I develop my conception of negative liberty further in Ch.4, I.iii.

integrity. Integrity has great power, but its power is *mysterious* as it both motivates public reason liberalism and threatens it with defeat. Integrity-based complaints seem both imminently relevant to assessing the ideal of public reason and to somehow misunderstand the concept of freedom upon which public reason liberalism rests. In forcing the public reason liberal to recognize the value of integrity, the integralist has introduced a free radical into liberal political theory. It is not clear what effect integrity will have. Consequently, the integralist cannot declare victory, nor must she accept defeat.

The point of this dissertation is to resolve the conflict between public reason liberals and their integralist critics. So what can we do? How do we master integrity's power? One strategy seems to be a non-starter. Suppose that we try to discern whether the costs public reason liberalism imposes on integrity exceed the benefits. But integrity is not the sort of value that can be balanced in this way. We cannot simply ask citizens to add up how much integrity they preserve if they submit to restraint and how much they preserve if they refuse. Integrity has a kind of deontic force, which is to say that sometimes it is objectionable to ask a citizen to sacrifice her integrity *even if* she would gain more of it in the end. Suppose the public reason liberal approaches a Christian with the following offer: burn the Bible and in the future you will have the ability to live a life of greater fidelity to your Christian principles. Even if there were a causal connection between Bible burning and living a more Christian life, the offer is unacceptable. Integrity requires not violating your principles and trading-off your projects and plans for greater gain. To have integrity in many cases is just to be the sort of person who is willing to sacrifice great value on behalf of her present commitments. It is for this reason that the citizen of faith cannot be satisfied with the offer of greater integrity in the future. To respect the citizen of faith, then, we must show her that public reason liberalism does not require her to sacrifice too much of her integrity *even if* she will gain more in the future. Integrity traded is integrity destroyed.

In one way, this was Rousseau's argument against Locke and Hobbes's versions of the social contract. Hobbes and Locke explicitly saw citizens as ceding the liberty of the state of nature in order to enter into human society. Hobbes claimed that one implication of the second law of nature was "that a man be willing, when others are so too ... to lay down this right to all things, and be contented with so much liberty against other men."⁸⁵ Locke claimed that in agreeing to the social contract, man "divests himself of his natural liberty, and puts on the bonds of civil society."⁸⁶ But

⁸⁵Hobbes 1994, p. 80.

⁸⁶ Locke 2003, p. 142.

Rousseau maintained that the dignity of the human person is diminished when she renounces her freedom: “To renounce liberty is to renounce being a man, to surrender the rights of humanity and even its duties.... Such a renunciation is incompatible with man’s nature.”⁸⁷ It is for this reason that public reason liberals seek a form of social and political life that leaves us “as free as before.”

For the same reason, integrity cannot be traded off with the other foundational values of public reason liberalism, such as respect for persons. Public reason liberals claim that to treat others as free and equal, to respect them, we must reason from their standpoint. To give up our authority to live by our own lights in order to engage in public justification is a kind of contradiction in terms for the public reason liberal. Integrity not only cannot override the requirement of respect for persons, it cannot *conflict* with it. If the two values are in tension, public reason liberalism contains a theoretical instability. Those who emphasize the importance of respect must therefore deemphasize the importance of integrity and vice versa. Yet both sources of normative concern seem of the utmost importance.

The public reason liberal is committed to showing that integrity and public justification do not conflict. If not, it seems impossible to explain to the citizen of faith why she should not appeal to her own private judgment. Rawls articulates this when he rhetorically exclaims: “Surely, the most fundamental questions should be settled by appealing to the most important truths, yet these may far transcend public reason!”⁸⁸ The question is this: how can we rationally choose between acting based on the totality of our personal judgments and acting based on reasons derived first and foremost from a public point of view? On the one hand, we always reason from our own standpoint, so it seems that we never truly escape reasoning based on our own private judgments. On the other hand, we expect each other to take account of the judgments of others. In a liberal society, we demand that citizens also recognize as significant the fact that honest and informed people disagree about what is right and good.

For this reason, the pressure to resolve the apparent tension between integrity and public reason is immense. And yet, insofar as public reason liberalism asks citizens of faith not to impose the totality of their values on others, it seems to heteronomously restrain their integrity. If restraint is heteronomous, the citizen of faith can still echo Rawls’s exclamation. Citizens of faith will cite their religious traditions as providing them with commitments of the greatest importance. Who can ask them not to act on what they believe are the answers? The citizen of faith is not obviously in the

⁸⁷ Rousseau 1997, p. 5.

⁸⁸ Rawls 2005, p. 215.

wrong to demand the authority to live by her own lights come what may. On the public reason liberal view, this is the most foundational and serious of all political claims. If the person of faith is sincerely convinced that her values are so sacred that she simply *must* shape her social world to fit them, the public reason liberal must be deeply concerned. For what can the public reason liberal say to her?

It may appear that the tension between public justification and integrity cannot be resolved. If private judgment never conflicted with public judgment, the problems that the public reason tradition developed to answer would never have arisen. Our private judgments diverge under conditions of freedom. As such, even if we mean well and understand one another's points of view, our aims will often conflict. Unless we permit the judgments of others to provide us with reasons for action that exclude our authority to act on our own private judgment, these conflicts cannot be rationally solved. All that can be achieved is a pragmatic peace, where one party simply exerts her judgment over others. This is the social state Rawls refers to as a *modus vivendi*.⁸⁹ The struggle of the public reason tradition, then, is to reconcile private and public judgment in order to preserve the freedom and dignity of man. To show that human beings remain free in civil society, the public reason liberal must show how it can be *rational* for an individual to permit public judgment to preempt her private judgment. Public reason liberalism claims that practical rationality requires a commitment to coercing others solely on the basis of the Public Justification Principle. Liberalism claims that those who only acknowledge the normative force of their private judgments undermine the very possibility of a liberal settlement. By following her conscience regardless of the judgments of others, the citizen of faith adopts a willingness to coopt others into her projects. But if practical reason requires taking the standpoint of others, then even the citizen of faith is committed to stepping outside of herself. In this way, public reason liberalism is compatible with her integrity.

VI: In Pursuit of a Strategy

But let us pause: Chapter 2 has only assessed the *mainstream conception* of public reason liberalism. We have yet to establish whether public reason liberalism *per se* conflicts with integrity. The integrity objection is leveled against the principles of restraint implied by the consensus conception of reasons. The promise of reconciliation lies in locating an *alternative conception* of public reason

⁸⁹ Ibid., p. xxxvii. The situation might also be worse than a *modus vivendi* in cases where there is no basic agreement on the governing principles of a society even for strategic reasons.

liberalism, one that reduces the apparent heteronomy of restraint so substantially that the remaining integralist complaints no longer seem pressing.

In my view, the reason there appears to be a tension between integrity and public reason is that liberalism is committed to recognizing a kind of *dualism about practical reason*. The power of public reason liberalism is that it recognizes that in politics there are two fundamental types of reasons for action: reasons derived from the standpoint of others and reasons to live lives of integrity. The claim of the liberal tradition and of all the great social contract theorists is that reasoning from one's personal standpoint alone is not only wrong but somehow *irrational*. The contract stories elaborated by early figures like Hobbes, Locke, Rousseau and Kant and later figures like Rawls, David Gauthier and others, are all attempts to show why individuals are committed by the structure of practical reason alone to establish institutions of public judgment. The irony is that liberalism, more than any other political tradition, is accused of selfish individualism. But it is *only* liberalism that requires reasoning from the standpoint of others even when personal value must be sacrificed. All other political traditions specify a vision of the good and ground reasons in bringing about that vision. Liberalism asks the individual to acknowledge the existence of reasons that originate in other points of view.

How then can we ask citizens to appeal only to a public conception of justice and not to the whole truth as they see it? Because in doing so, we ask citizens to live a rational life. The dual structure of liberal practical reason shows that practical rationality itself requires restraining integrity and thus restraining integrity is no violation of freedom. The solution to the religion and politics problem is therefore to show how practical rationality can contain two types of reasons, some of which derive from our own perspective and those that derive from the perspectives of others. If liberal practical reason indeed has a dual structure, it will sometimes be rational to restrain one's integrity. If so, the integralist will be rationally committed to paying some of the costs about which she complains. In order to illuminate this dual structure, however, we must go beyond political theory into moral theory. For we need a plausible model of practical reasoning to make sense of this tension within the liberal tradition. I find contractualism provides an attractive model, but only a specific version. By looking for an attractive version of contractualism, we may discover a method of resolving the apparent tension within the structure of liberal practical reason. If so, we can resolve the religion and politics debate.

Chapter 3: A Model of Liberal Practical Reason

Chapter 2 ended with a promise: model the structure of liberal practical reason and a strategy for reconciling public reason liberals and integralists will emerge. Chapter 3 is an attempt to produce such a model. Public reason liberals and integralists point to two distinct types of reasons, reasons derived from the integrity of citizens and reasons derived from a public point of view. Remember the problem developed in Chapters 1 and 2: these two sorts of reasons do not bear obvious relations to one another. Public reason liberals are frequently unmoved by integrity objections because they think that reasons derived from a public point of view are authoritative, but integralists maintain that reasons of integrity are authoritative. In other words, both think their preferred sort of reason is the winner in a fight. My project is to assume that both sorts of reasons have normative force and so I must explain how one sort of reason can trump the other, or when it is appropriate to act on reasons of integrity or reasons derived from a public point of view. It is for this reason that I must model this “dualist” conception of practical reason. I must explain how these two types of reasons are related to one another and, prior to this, I must say what these two types of reasons *are*.

Yet political theory provides few resources for giving a precise account of these types of reasons and their relations to one another. I will turn to moral theory for help. I take the liberal “dualist” model of reasons to comport well with a deontological conception of practical reason which acknowledges as basic reasons rooted in respect for persons while simultaneously acknowledging that there are goal-based or telic reasons. Deontologists typically resist rooting the normative force of deontic reasons, or reasons to respect persons, in goal-based or telic reasons. For this reason, they are committed to a form of dualism about practical reason similar to the form of dualism that to which I believe the liberal political tradition is committed.

The contractualist tradition is an especially attractive area of moral philosophy from which to draw. Public reason liberalism requires that justificatory reasons be those that citizens can reasonably be expected to recognize, and contractualists hold that the normative force of many reasons is rooted in some form of reasonable, mutual recognition. Both traditions maintain that these reasons are frequently authoritative for those who have them and share an internal recognition condition. The contractualist tradition also helps provide resources for building a dualist model of liberal practical reason because it roots the force of deontic reasons in respect for persons, just as public reason liberalism grounds the requirement that coercion be publicly justified in respect for persons.

Finally, contractualist moral theories are frequently conceptually richer than conceptions of public reason employed in the religion and politics debate.

For these reasons I will review the structure of contractualism and identify the relevant elements of contractualism for my project. Unfortunately, it is hard to explain what an account of liberal practical reason needs in the abstract. Instead, I will propose an appropriate model of practical reason in Chapter 4 and contrast it with the mainstream version of contractualism I criticize in Chapter 3. For this reason, Chapter 3 advances a set of criticisms of mainstream contractualism and Chapter 4 both develops an alternative and shows how it can solve the conflict between public reason liberals and integralists.

I must be clear that while the account I will develop presumes the validity of some contractualist ideas, my aim is not to vindicate contractualism. Instead, I employ contractualist resources in the hopes of explaining how reasons of integrity and reasons derived from a public point of view relate to one another. Public reason liberals need not be contractualists in moral theory or vice versa. Thus, while contractualism helps to articulate an attractive version of public reason liberalism, I draw on contractualism merely to help explicate public reason liberalism's deep structure.

Contractualist moral theory currently stands under the shadows of two giants: Thomas Scanlon and Stephen Darwall. Their sophisticated versions of contractualism therefore serve as a natural starting point for developing a model of liberal practical reason. To build my own approach to contractualism without taking their views as points of departure would be inappropriate. There is no need to start from scratch. Nonetheless, while I begin with their accounts, the accounts require modification. In this chapter, I lodge two criticisms against Scanlon-Darwall mainstream contractualism: (i) mainstream contractualism draws an implausible connection between human dignity and a duty to publicly justify interference, moral demands and the like, and (ii) mainstream contractualism assumes an excessively dyadic account of public justification that obscures the role of social rules in large-scale modern societies. In Section I, I introduce the mainstream approaches of Scanlon and Darwall. Section II advances the first criticism, whereas Section III advances the second criticism. Section IV concludes. In explaining the weaknesses of mainstream contractualism, I lay the groundwork for the model of liberal practical reason presented in Chapter 4.

I: Mainstream Contractualism

I.i Contractualism Defined

Contractualism identifies justifiability to others as the core of interpersonal morality. Thomas Scanlon defines contractualism as follows:

An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement.¹

The contractualist formula purports to explain the rightness of actions and the authority of the moral demands associated with them. Acts are wrong if the principles they violate cannot be reasonably rejected. When a moral principle cannot be reasonably rejected, the moral demands derived from it are authoritative. For contractualists, moral demands must be justified in part because moral demands are impositions. By demanding that Reba act rightly, John insists that Reba act as she might otherwise not regardless of whether she would like to do so. John arguably imposes upon Reba in two ways. First, John interrupts Reba's ordinary activities and she must respond by changing her behavior in some way or another. But contractualists often maintain that Reba has reason to care about John's *moral* demands, that human beings care about being in moral relations with one another, relations of equality and without blame or sanction. Scanlon, for instance, has maintained that people are naturally motivated to act in ways that are justifiable to others.² Ostracism and blame originating from one's moral community is therefore undesirable. Consequently, Reba will feel pressure to conform to justified moral principles, and so John's moral demands impose upon Reba in another manner as well. Contractualists also argue that our actions must be justifiable to others and that actions which cannot be justified demonstrate a failure to appreciate or respond appropriately to the dignity or value of persons. These are merely the basics. The particulars of Darwallian and Scanlonian contractualisms are examined in I.ii and I.iii respectively.

¹ Scanlon 1998, p. 153.

² Ibid., pp. 153-158.

I.ii Darwallian Contractualism

Darwall grounds contractualism in the “second-person standpoint”:

My aim ... is to argue that what is common to contractualist approaches, that we are mutually accountable for regulating ourselves by principles that are acceptable (in some suitable sense) to each as equal moral persons, can itself be grounded in, and is best appreciated from the second-person standpoint. Or, to put the point the other way round: the most promising way to work out the content of moral principles we are committed to by the equal dignity we must presuppose in the second-person perspective is through some version of contractualism.³

Right and wrong are grounded in “the perspective you and I take up when we make and acknowledge claims on one another’s conduct and will.”⁴ This standpoint “commits us to regulating our conduct by principles that are acceptable, or not reasonably rejectable, to each as free and rational agents.”⁵ The second-person standpoint can only be defined by a “circle of irreducibly second-personal concepts.”⁶ The circle has six elements: (a) standpoints, (b) second-person reasons and authority, (c) responsibility, (d) valid claims and demands, (e) members in a community of moral agents, and (f) the idea of equal dignity.⁷

(a) What gives the second-person standpoint its “second-personality”? That is, when we take the second-person standpoint, what exactly are we doing? For Darwall, we take the second-person standpoint when we engage in moral practices that are essentially *interpersonal* and recognize that the force of these practices cannot be reduced to a mere first-personal point of view.⁸ For Darwall, duties are generated through these social practices, as are our reasons to act upon them. To take the

³ Darwall 2006, p. 304.

⁴ Ibid., p. 3. I take it our practices of making and acknowledging claims on each other’s conduct and will is tantamount to engaging in the practice of interpersonal morality.

⁵ Ibid., p. 300.

⁶ Ibid., p. 11. The key notions ... “comprise an interdefinable circle; each implies all the rest. Moreover ... there is no way to break into this circle from outside it.” Ibid., p. 12.

⁷ Ibid., p. 12. Darwall’s view contains other concepts that I will not discuss in detail, in particular, autonomy and rationality. His view can be explicated with a smaller number of concepts, however. To streamline the discussion, some concepts will not receive our full attention. Note that by ignoring autonomy, I ignore only Darwall’s notion of autonomy, not the positive liberty view I discussed in Chapter 1.

⁸ And while I will not discuss a third-personal or agent-neutral perspective, it too is ruled out of accounting for interpersonal morality on Darwall’s view. When we take a second-person point of view, we also avoid impersonal, third-person points of view as well.

second-person standpoint is to consider moral claims from a common point of view and acknowledge that the justification of our moral demands on others depends upon their reasons and not ours alone. This is because we see *other people* as morally related to us, as standing in relations of accountability. In this way, the second-personal standpoint embodies a certain conception of reciprocity since we recognize that the generation of duties depends on some sort of convergence between our reasons to make demands and others' reason to acknowledge them. What's more, we also recognize that this social manner of generating moral duties applies to all in precisely the same way. The duty to justify our moral demands based in part on the reasons of others applies as much to ourselves as to others.

(b) The second-person standpoint generates duties and justifies moral demands via a normative element known as a second-person *reason*. Second-person reasons obligate others and do so in part because their very soundness as reasons “depends on presupposed authority and accountability relations between persons and, therefore, on the possibility of the reason’s being addressed person-to-person.”⁹ Second-person reasons are employed to justify interpersonal claims and our practices of praising and blaming others. Darwall argues that by offering second-person reasons, we presuppose that we have a claim or *standing* to hold others accountable for violating or not responding to our demands. Thus if John has the appropriate standing, community members recognize his authority to hold others accountable and their same power over him. Community members exercise their authority over one another by issuing second-person reasons, since such reasons hold others to moral norms they already regard as having moral force. Without presupposed authority and accountability relations, second-personal reasons cannot generate obligations. An essential feature of second-personal reasons is that they are essentially “agent-relative” because they make a “non-trivial and uneliminable pronominal back-reference” to the person who has the reason in question.¹⁰ This means that agent-relative reasons are reasons that are essentially “reasons for” a particular agent. The agent-relativity of second-personal reasons generates their recognitional component. For John to obligate Reba, she must recognize his demand as having moral force. She must do so through her appreciation of reasons that apply to her, or reasons that are relative to her. Unless these reasons can be so recognized, they fail to obligate. Second-personal reasons are thus agent-relative reasons that have the power obligate those to whom they are addressed.

⁹ Ibid., p. 8.

¹⁰ Ridge 2003, p. 339.

(c) Darwall relates responsibility and accountability in the second-person standpoint to P.F. Strawson's conception of the connection between responsibility and normativity. Darwall terms "Strawson's Point" the claim that "desirability is a reason of the wrong kind to warrant the attitudes and actions in which holding someone responsible consists in its own terms."¹¹ According to Darwall, Strawson claims that we can only hold others morally responsible for their wrongful actions when they recognize a reason not to act wrongly. The attitudes of praise and blame associated with holding others to account are "reactive attitudes" that consist in "an expectation of, and demand for" behaviors from others in a particular context.¹² Since reactive attitudes concern what claims we can hold others to, Darwall holds that they "presuppose the authority to demand and hold one another responsible for compliance with moral obligations."¹³ We cannot obligate others merely by willing that we do so, merely by desiring that the duty obtain. Instead, our social practices and attitudes commit us to addressing the reasoning of others.

(d) To make a valid claim or demand on the conduct and will of others, several preconditions must be met. First, those making claims must take the second-person standpoint and offer second-personal reasons. They are required to give reasons to others which those individuals cannot reasonably reject. But claim-makers acquire their second-personal authority by acknowledging—at least implicitly—a claim of reciprocity, that others can make claims on our conduct and will just as we can make the same claims on theirs. As Darwall argues,

There is thus, I argue, a form of reciprocal respect that is built into all second-personal reason-giving.... Any pure case of claiming or demanding presupposes the standing necessary to enter into second-personal reasoning at all. Specifically, it presupposes a distinction between legitimate forms of address.¹⁴

Darwall follows Fichte in distinguishing between legitimate forms of address that "determine themselves freely by second-personal reasons" and illegitimate forms of address that merely coerce.¹⁵ An appropriate form of address is one that can obligate and only second-personal forms of address are appropriate to making valid claims and demands on others. To give an example, our discussion

¹¹ Darwall 2006, p. 15. For the original essay, see Strawson 1974, pp. 1-28.

¹² Strawson 1974, p. 15.

¹³ Darwall 2006, p. 17.

¹⁴ *Ibid.*, p. 21.

¹⁵ *Ibid.* Also see Fichte 2000, p. 41.

of Locke on the logic of persecutors provides an illuminating example of an inappropriate form of address. Suppose John demands that Reba, a member of another church, not practice her religion openly. The reason John provides Reba is that he affirms the true religion and that Reba does not. When Locke reminds his reader that *every church is orthodox to itself* he is arguing that John's claim that he affirms the true religion is the wrong kind of address.¹⁶ His reason is not second-personal because it is not one Reba recognizes as normative, for she does not believe John has the true religion. What's more, the claim fails to acknowledge reciprocal relations of second-personal authority, as John would not acknowledge an equivalent claim on his conduct by Reba. As such, John's form of address is inappropriate and cannot generate a valid claim.

(e) To make justified demands on others, we must treat them as members of a community of equals. Members of a community of equals acknowledge relations of reciprocity and moral justification. They presuppose a kind of standing, "the standing necessary to enter into second-personal reasoning at all."¹⁷ For Darwall, this standing derives from membership in a single human community that implicitly acknowledges the universal authority of persons to make bona fide second-personal claims on each other's conduct and will. Further, community members cannot make legitimate claims on others without granting the same authority to others: "Whatever more specific difference-defining authorities they presuppose, each is committed also to assuming that free and rational persons have a dignity or authority in common to address (and be addressed) by second-person reasons."¹⁸

(f) One of Darwall's central claims is that "that which is called the inviolable value or *dignity* of persons has an irreducibly second-personal element, which includes the authority to demand certain treatment of each other...."¹⁹ That humans have equal dignity explains how they can be both the objects and subjects of moral demands. When John imposes a justified moral demand on Reba, he issues a second-personal reason. He is then permitted to hold Reba accountable for not satisfying the expectation set by the reason. Yet it is only because persons have equal dignity that John must address Reba in terms that she can recognize. Arbitrarily asserting his moral reasoning over Reba's is incompatible with John respecting her dignity. To respect persons, we must admit what Darwall calls "Fichte's Point," that second-personal claims presuppose "a common competence" to recognize valid claims and that all claims that we have reasons to act presuppose "the capacity to determine

¹⁶ Locke 2003, p. 225.

¹⁷ Darwall 2006, p. 21.

¹⁸ Ibid.

¹⁹ Ibid., p. 13.

[ourselves] by these reasons.”²⁰ The fact that human beings share a dignity means that we must respect this capacity.

Darwall seems to suppose that a person’s dignity by itself creates moral duties. Specifically, the fact that each person has a dignity *just is* what makes them subjects of justification. It is in virtue of our dignity that we have the right to require that others impose justified moral demands upon us. Again, dignity entails that no one can exercise legitimate moral authority unless their claims can be justified in second-person terms. Further, Darwall appears to take the fact that humans have dignity to *by itself* imply the standard of justifiability to others. Darwall argues that the irreducibly second-personal element of dignity “*includes* the authority to demand certain treatment of each other.”²¹ Similarly, the idea of dignity “partly *involves* there being ways one must conduct oneself toward persons and ways one must not ... persons are beings who may not be treated in certain ways.”²² It is as if dignity partly *consists in* possessing the power to impose duties and obligate others. Dignity entails that persons have moral obligations to justify themselves to others because each person has an authority to demand a certain sort of treatment, that is to be treated as a subject of justification.

I.iii Scanlonian Contractualism

Darwall intends his work to provide a foundation for contractualist moral theory. In this way, he intends his framework to clarify the Scanlonian contractualism to which we now turn. In addition to justifiability to others, Scanlon’s contractualism possesses three main conceptual components: (a) standpoints, (b) reasonable rejection, and (c) demanding treatment.

(a) Like Darwall, Scanlon argues that “we have reason to consider whether there are standpoints other than our own present standpoint from which the principles we are considering could reasonably be rejected.”²³ Moral justification requires taking others’ point of view in two senses, first as “those to whom justification is owed” and second as “those who might or might not be able reasonably to reject certain principles.”²⁴ Taking the standpoint of others not only involves considering the effect proposed principles will have on others but also how they will evaluate those

²⁰ Ibid., p. 21.

²¹ Ibid., p. 13. Emphasis mine.

²² Ibid. Emphasis mine. And again: “Our dignity as persons *includes* [emphasis mine] ... an irreducibly second-personal authority to demand respect for this very authority and for the requirements with which it gives us the standing to demand compliance.” Ibid., p. 14.

²³ Scanlon 1998, p. 202.

²⁴ Ibid.

effects. Justifying a principle requires considering the reasons others recognize. Thus, standpoints figure into contractualism because they are points of view of other reasoners. Scanlonian standpoints are in this way analogous to the Lockean idea. To make justified demands of others, we cannot merely consult our own private judgment about what is justified and true. Instead, we must appeal to common standards of reasoning that others recognize as valid. For Scanlon, deciding whether a principle is justified requires considering it “from a number of standpoints.”²⁵

(b) Scanlon requires that moral justification locate principles that cannot be *reasonably rejected*. The concept of the reasonable is nuanced. First, it “presupposes a certain body of information and a certain range of reasons which are taken to be relevant and goes on to make a claim about what these reasons, properly understood, in fact support.”²⁶ A person is reasonable only when she is sensitive to the non-moral information and reasons appropriate to the moral determination at hand. Second, being reasonable is distinct from being rational. It essentially contains “moral content”.²⁷ To determine what is reasonable, we must weigh objections, specifically “objections to permission” and “objections to prohibition.”²⁸ Objections to prohibition are reasons to reject the proposed principle. Determining which acts are reasonable requires appealing to principles that are not themselves hard and fast rules, but “general conclusions about the status of various kinds of reason for action.”²⁹ While principles “may rule out some actions” they “leave wide room for interpretation and judgment.”³⁰ Deciding what is reasonably rejectable therefore requires considering (i) which reasons we acknowledge as having normative force, (ii) which reasons others acknowledge as having normative force, (iii) the weight and number of the reasons, (iv) which principles represent or account for these reasons and (v) which principles are reasonable. We may then and only then determine which principles can be reasonably rejected. Standpoints, principles, and reasonableness fit together to determine which norms are morally justified. Rejectability is the second component of reasonable rejection. If a reasonable person sees herself as having reason to reject a proposed moral principle, then the proposed principle is unjustified. A principle is rejectable just when it is unjustified. Providing a model of what the reasonable person will reject explains what principles she in fact has reason to obey.

²⁵ Ibid., p. 213.

²⁶ Ibid., p. 192.

²⁷ Ibid., p. 194.

²⁸ Ibid., p. 195-6. Note here that Scanlon rejects what I call the symmetry requirement on public reasons. See Ch.5, IV.

²⁹ Ibid., p. 199.

³⁰ Ibid.

(c) Contractualism is grounded in the idea of a demand for treatment. If principles of social morality are justified only when they cannot be reasonably rejected, then all individuals have the claim-right to demand that they not be imposed upon based on principles they can reasonably reject. For Scanlon, "... proper respect for [persons'] distinctive value involves treating them only in ways that they could, by proper exercise of this capacity, recognize as justifiable."³¹ If John respects Reba, then he acknowledges her right to demand a certain sort of treatment. Like Darwall, the idea of respect for persons or the dignity of persons simply entails that individuals have a right to demand treatment from others.

Darwall uses the second-person standpoint to ground several of Scanlon's key ideas. First, the notion of the reasonable can be vindicated "by presupposing the equal dignity of persons" which thereby "commits us to contractualism's deepest idea: a community of mutually accountable free and rational persons."³² Reasonable people engage in "reciprocal recognition" of others and issue reasons compatible with that recognition of equal dignity.³³ A reasonable demand is one that rests on this form of recognition. Otherwise, the demand will be rooted in reasons of the wrong kind, running afoul of Strawson's Point. Reasons of the wrong kind are reasons that do not appeal to the perspective of others, that do not respond to the reactive attitudes of others and do not recognize their status as having dignity. Without such reasons, those who make demands cannot do so authoritatively. In other words, the wrong kind of reasons cannot obligate. Accordingly, reasonable rejection becomes the criterion of admissible second-person reasons because the standard indicates which reasons others recognize as having normative force. Darwall's deepening of Scanlon's idea of standpoints should already be clear from the explication above.

We now have a workably clear notion of mainstream contractualism. With these resources outlined in some detail, I can now turn to raise objections to some of their elements with the aim of preserving other elements in Chapter 4. Remember that I am simply laying out conceptual resources from which I will pick and choose for our purposes. Seeing some problems with the connections among contractualist ideas will help explain which ideas we will use and which we will modify.

³¹ Ibid., p. 169.

³² Ibid., p. 320.

³³ Ibid., p. 318.

II: Dignity and Moral Justification

II.i Scanlon, Darwall and Dignity

Mainstream (Scanlon-Darwall) contractualism takes the idea of dignity to imply that all persons are under a natural duty to treat others as subjects of justification. Consequently, when John makes a normative demand of Reba, he has a duty to issue demands that can be justified to Reba. Otherwise, he fails to respect her dignity. She has the authority to demand that John justify his moral demands on her in terms she can reasonably be expected to accept. For Scanlon and Darwall, the fact that all persons merit respect straightforwardly implies that everyone has a right to demand that they be treated in accord with principles they cannot reasonably reject. In this way, Scanlon and Darwall can be understood to affirm a conceptual implication from dignity to the right to demand being treated as a subject of moral justification. All moral theories take certain conceptual claims as basic in the sense that some moral ideas are not justified in moral foundational terms. For Scanlon and Darwall, the idea of dignity seems basic. Persons are subjects of justification in virtue of their status as persons. The duty to treat others as subjects of justification is not itself justified in simpler terms.

Notice that Scanlon and Darwall offer no argument for this conceptual connection.³⁴ If they do not take the move from dignity to the right of persons to demand justification as basic, then it is unclear what their argument is. For this reason, it seems most charitable to read them as taking this claim as basic. Consequently, criticizing this conceptual claim does not require refuting their positive arguments but rather showing that their conceptual claim is vulnerable to a straightforward counterexample which demonstrates a conceptual gap between dignity and moral justification.

II.ii The Gap between Dignity and Moral Justification

While all moral theories must take some normative concepts as basic, Scanlon and Darwall's choice to take dignity as basic is implausible. Dignity or respect for persons does not carry enough

³⁴ Scanlon does gesture towards an argument when he claims that the contractualist ideal characterizes "the relation with others the value and appeal of which underlies our reasons to do what morality requires. This relation, much less personal than friendship, might be called a relation of mutual recognition. Standing in this relation to others is appealing in itself—worth seeking for its own sake." Thus the reason to embrace justifiability to others is that it is characteristic of a relationship that we believe has some sort of intrinsic value. The problem with this suggestion is that it seems to repeat the conceptual claim in another form, that is, that the dignity or value of persons means that there is something valuable in standing in the relation with them that is achieved when our norms are justifiable to others. See *ibid.*, p. 162.

conceptual content to imply that others have duties to make only justified demands on others.³⁵ Accordingly, the claim that the dignity of persons entails a duty to issue justified demands must be scrutinized.³⁶ In this spirit, let's examine how dignity and moral justification might come apart. The Scanlon-Darwall natural duty to make justified demands is directly tied to the notion of public justification. This natural duty is, in effect, a duty to issue publicly justified moral demands. Since Scanlon and Darwall are contractualists, they believe that moral claims must be justified according to reasons that one cannot reasonably reject, which is tantamount to affirming a duty to issue only publicly justified moral demands. If we can drive a wedge between dignity and the requirement of public justification then we will have *mutatis mutandis* driven a wedge between dignity and a contractualist conception of moral justification.

Christopher Eberle's attempt to drive a conceptual wedge between respect for persons and the public justification of coercion can be used to bring out the conceptual gap between respect for persons and the public justification of moral norms. Eberle acknowledges that there is an intuitive link between respect for persons and public justification of coercion but nonetheless believes that compelling counterexamples can sever the bond. Consider Eberle's case of Elijah, a Christian so concerned about material inequality that he supports a radical redistribution of wealth. Elijah only supports redistribution because he embraces Liberation Theology, which holds that God has a preference for the poor. This divine preference, Liberation Theologians claim, requires us "to take quite drastic measures in narrowing the disparities in life opportunities between rich and poor."³⁷ Eberle thinks that Elijah should *pursue* public justification but that if he fails to publicly justify his support for radical redistribution, he may still permissibly coerce others into redistributing their wealth. In fact, Eberle's argument against public reason liberalism generally is that the arguments used in its defense only mandate the *pursuit* of public justification rather than the achievement of public justification. For this reason, citizens must only attempt to make justified demands, but they are under no duty to refrain from engaging in coercive action should they fail after an honest attempt. Again, for Eberle, the intuitions that are marshaled to connect respect for persons and public justification do not unambiguously support the achievement of public justification over its

³⁵ Throughout the dissertation I take the notion that humans have dignity to conceptually imply that they are owed respect. Thus I will use "dignity" and "respect for persons" as synonymous.

³⁶ We might think that Darwall and Scanlon take dignity to be embodied by having the standing vis-à-vis others to demand justification, but this is, again, to simply make the same conceptual claim in another form.

³⁷ Eberle 2002, p. 112. Eberle also stipulates that Elijah's evidential set gives him a high degree of rational justification for his beliefs.

pursuit. Eberle looks to arguments offered by public reason liberals to resolve this ambiguity and claims that no such arguments are to be found.³⁸

Let us transfer Eberle's distinction between pursuing and achieving public justification for coercive laws to pursuing and achieving the public justification of moral norms. Eberle argues that respect for persons requires pursuing public justification. I assume that Darwall and Scanlon will agree with him on this much. But they will part ways concerning whether the achievement of public justification is required to respect persons. Suppose, for instance, that Elijah wants to hold others accountable for not wanting to redistribute wealth, but not coerce them. That is, he is content simply to blame and ostracize his compatriots for not supporting his preferred proposals rather than coercing them. Elijah tries earnestly to convince his compatriots that a norm requiring radical redistribution is justified and that they should comply with it. But suppose that they continue to disagree after lengthy discussion. The process of interpersonal justification proceeds just as Scanlon and Darwall would have it but the process does not result in a mutually recognized justification of the norm. Elijah continues to think that his compatriots are dead wrong, whereas they are convinced that they do no wrong by refusing to redistribute as they do. What is Elijah to think? Should he still attempt to interfere with them through blame and ostracism? Should he hold them accountable for failing to see things his way? If Scanlon and Darwall want to say no, then what should they claim in reply? They will assuredly want to maintain that in so doing Elijah disrespects his compatriots and fails to take their dignity seriously. But suppose that Elijah remains sincere. He is convinced, say, that his compatriots are simply being difficult or have failed to appreciate the evidence due to impatience, confusion or mere moral and spiritual blindness. Elijah believes that the imperative to redistribute is God's will and so he assumes that his compatriots are simply confused. Darwall and Scanlon think there is nothing more to be said. Elijah has tried and if he respects his compatriots, he will stop. He should conclude that the norm is not justified to them and out of respect cease harassing or browbeating them. Scanlon and Darwall will claim that Elijah's compatriots have the authority to demand that he treat them with respect and thus they have the authority to demand from him a justification for his continued action. But why should Elijah acknowledge such a duty? I will argue in Chapter 4 that the best explanation for why Elijah should stop is that he has failed to satisfy a presumption against interference that is implied by our shared conceptions of ourselves as autarchic or self-directing (Ch.4, I), but it seems to me that Darwall and Scanlon cannot get away with saying only what they have so far.

³⁸ Ibid., p. 115.

I want to emphasize that I believe that Darwall and Scanlon can give an attractive reply but that it will look like the Stanley Benn and Gerald Gaus inspired presumption against interference I defend in Chapter 4. Eberle's distinction seems to show that for now Darwall and Scanlon have failed to establish the tie between respect for persons and public justification. At a minimum, Scanlon and Darwall must say more than they have. I prefer a version of contractualism that eschews taking the conceptual connection between dignity and demanding treatment as basic.³⁹ To claim that respect for persons means responding to their right to demand a certain sort of treatment is ambiguous. Instead of explicating the key moral ideas that connect respect for persons and contractualism, Scanlon and Darwall leave us with a mere intuitive link. Eberle's analysis shows that an intuitive link is not enough.

III: Two Functions of Social Morality

III.i Scanlon, Darwall and Dyads

Scanlon and Darwall often represent moral justification and enforcement as set in dialogue between two individuals. Moral justification is analogized either with a discussion about whether a moral rule is justified or the enforcement of a rule between one who violates the rule and the one who blames him for doing so. Hence public justification is frequently depicted as occurring between a *dyad* of persons. Darwall especially places emphasis on "second-personal address" and imagines the moral community's enforcement of moral rules as occurring in the course of day-to-day moral dialogue. In recent work, Scanlon has argued that blaming others consists in taking the blameworthy action as indicating "something about the person that impairs one's relationship with him or her and to understand that relationship in a way that reflects this impairment."⁴⁰ The practice of blaming someone for violating a justified moral norm requires a "relationship". Scanlon extends his account of blame to the social world by suggesting that we have a "moral relationship" with all of humanity.⁴¹ In mainstream contractualism, moral justification seems to occur pair by pair.

Scanlon and Darwall likely do not regard their theories as dyadic. Instead, dyadic conceptions of justification serve merely as hypothetical simplifications of complex social processes.

³⁹ I defend the presumption against interference in Ch.4, I. For Benn's defense, see [BENN ENDNOTE], pp. 87-90. For Gaus's defense, see Gaus 1990, pp. 379-395.

⁴⁰ Scanlon 2008, pp. 122-3.

⁴¹ *Ibid.*, p. 139. Scanlon recognizes that many will see his account as a stretch. See *ibid.*, pp. 146, 175.

And yet Scanlon and Darwall do not extend their analyses to problems that arise at the N-person level of moral justification. I shall give two arguments that relying on the dyadic analogy is problematic. First, the dyadic analogy obscures the difficulties in communicating reasons to large groups. Specifically, the dyadic analogy renders obscure how large groups of often anonymous individuals could ever converge on publicly justified moral norms. In most justificatory contexts—even plausible hypothetical ones—contractualist justification cannot take place dyad by dyad. In fact, most of those involved in moral justification will likely never be aware of each other’s existence. Second, the dyadic model obscures distinctions between the diversity of reasons relevant to publicly justifying moral norms. Articulating a satisfying account of public reasons and second-person reasons in particular requires grasping the diversity of types of justificatory reasons.

While Scanlon claims that his account of contractualist justification is merely hypothetical, Darwall frequently discusses real-world cases of second-personal address and moral conversation.⁴² In this way, Darwall’s theory attempts to justify moral authority as it is practiced. Should Scanlon embrace a purely hypothetical contractualism, one that characterizes a class of moral principles and explain why they are moral, he can avoid the problems I raise below. Scanlon’s theory need not explain how persons acquire moral authority over others through a real-world process, but by eschewing this aim, Scanlon will give up one of the most attractive features of contractualist moral theory. Darwall’s question about how our actual practices of making moral demands are justified cannot be answered with an account of hypothetical justification. Accordingly, our first argument focuses on Darwall.

III.ii Moral Rules

Before criticizing dyadic accounts of justification, we need an account of moral rules. A moral rule specifies a class of moral actions or moral demands. For example, a moral rule that requires keeping one’s promises describes a class of actions—promise-keeping—and requires that each promise is kept. Moral address is usually made by means of moral rules. In other words, moral rules *mediate* second-person address. If John blames Reba for not keeping her promise, he blames her for not

⁴² Scanlon 1998, p. 155. This account of hypothetical justification is not to be confused with Scanlon’s views on well-being as he rejects an informed desire account. See p. 55.

following the promise-keeping rule.⁴³ Norms of behavior must satisfy several conceptual conditions to count as moral rules. Kurt Baier argues that there are six. Moral rules are:

- (i) Part of the mores of the group;
- (ii) Supported by the characteristically moral pressure;
- (iii) Universally teachable and therefore universalizable;
- (iv) Not merely a taboo;
- (v) Applied in accordance with certain principles of exception and modification;
- (vi) Applied in accordance with certain principles of application.⁴⁴

Baier's conception of moral rules fits well with the account of public justification developed in Chapter 1. First, Baier maintains that moral rules must be recognized within a community. This is to say that they must be generally acted upon and recognized as binding. Rules must also be public by being known to be rules of the group. This is what is meant by a "more" of the group in question. Second, Baier's idea of "characteristically moral pressure" is similar to using blame as a method of enforcing social morality. All contractualists agree with Baier that moral rules must be universal.⁴⁵ I conceive of universalizability as restricted in scope. The "universal" application of some moral rules might apply only to community members or to the group to whom justification is owed. As P.F. Strawson notes, moral rules need not have "the character of applications of universal principles holding for all men."⁴⁶ Universality is defined as the application of rules to all members of a group. For Baier, universality bars counting three types of social rules as moral: (i) "self-frustrating" rules, or ones that by nature cannot be adopted by all members of a group, (ii) "self-defeating" rules, or ones defeated as soon as "its adoption by someone is revealed by him ..." and (iii) rules that are "literally impossible to teach."⁴⁷ Moral rules are essentially reversible in that they apply to all persons when they share similar characteristics; when John demands that Reba not violate a rule, he must also regard himself as subject to the rule. As Strawson notes, universalizability points to "something genuinely universal in morality: the necessary acceptance of reciprocity of claim."⁴⁸ By "reciprocity of claim" Strawson means something like Baier's idea of reversibility. Darwall acknowledges this

⁴³ These rules can still be second-personal if they enjoy interpersonal recognition, as Baier-Strawson moral rules do.

⁴⁴ Baier 1954, p. 108.

⁴⁵ For Scanlon's account of generality, see Scanlon 1998, pp. 206-213.

⁴⁶ Strawson 1974, p. 41.

⁴⁷ Baier 1954, p. 111-2.

⁴⁸ Strawson 1974, p. 41.

point as well, though he notably does not apply it to moral rules.⁴⁹ Baier also argues that moral rules are distinct from taboos because they are capable of being true or false; in our terms, moral rules admit of being justified or unjustified. Finally, moral rules are applied in line with principles of exception, modification and application just as publicly justified norms can be modified and applied in line with the public justification.⁵⁰

The foregoing is a description of positive moral rules. They become normative—for the public reason liberal and the contractualist—when they are publicly justified. To be compatible with the freedom and equality of persons, moral rules must permit sanction when they are violated. For interference involved in sanctioning to be justified, each person must have a conclusive reason to endorse the rule upon whose violation the sanction is based. We may now reformulate the Public Justification Principle to apply to moral rules:

The Public Justification Principle: A coercive action C is justified iff it is a constituent of rule M that members of the public P have conclusive reason R to endorse.

Since moral rules refer to classes of actions, they may vary widely in scope. A moral rule might be a “mid-level” rule in that it prescribes a specific set of actions. “Say ‘Bless you’ when someone else sneezes” is an example. It requires saying “Bless you” for any particular sneeze. In contrast, a rule might be more general, like “Do not steal.”

For Baier, like most moral philosophers, moral rules must be *overriding*. In other words, they must provide reasons for action that outrank reasons to act on other considerations. Baier notes that “[i]t is generally believed that when reasons of self-interest conflict with moral reasons, then moral reasons override those of self-interest. It is our common conviction that moral reasons are superior to all others.”⁵¹ When self-interested reasons to act conflict with moral reasons, moral reasons trump. For reasons of space, I will take for granted that moral rules are overriding even over many non-self-interested reasons. In short, I regard moral rules as overriding because they cannot perform their coordinating function unless they are so regarded. The liberal tradition has always emphasized

⁴⁹ Darwall 2006, p. 23.

⁵⁰ I explain how to identify justified moral rules in Chapter 10.

⁵¹ Baier 1958, p. 148. Note that Baier does not always restrict his conception of self-interest to non-altruistic conceptions of interests and value. In later work, Baier understands moral rules as trumping all “self-anchored” reasons. Baier 1995, p. 72.

that without a common political morality of rules with overriding status, social cooperation cannot occur or is much less effective than it would otherwise be. Putting it in Baier's terms:

It will be clear to everyone that universal obedience to certain rules overriding self-interest would produce a state of affairs which serves everyone's interest much better than his unaided pursuit of it in a state where everyone does the same. ... The very *raison d'être* of a morality is to yield reasons which override the reasons of self-interest in those cases where everyone's following self-interest would be harmful to everyone. Hence moral reasons are superior to all others.⁵²

Social cooperation is only possible with moral rules and moral rules must be regarded as overriding to play this coordinating role. Otherwise, people would violate moral rules when they proved too inconvenient to an individual's pursuit of personal value.⁵³ The system of moral rules would collapse and persons in a society would no longer have the assurance that they could cooperate for their mutual gain.⁵⁴

III.iii Moral Rules and N-Person Public Justification

An extended social order contains millions of individuals cooperating in order to pursue their diverse ends. A justified liberal social morality, therefore, is composed of rules that facilitate cooperation on a staggering scale. Yet since the social order is so complicated, information about why rules are publicly justified is almost always unavailable to others. As economist and social theorist F. A. Hayek observed, knowledge of economic and social information is inherently dispersed. Hayek is known for arguing against the feasibility of central planning. Many think his challenge to central planning is that central planners cannot engage in the calculations necessary to generate economically efficient plans once the relevant information is compiled. However, Hayek argues that the primary problem is *collecting* the relevant data since "... the 'data' from which the economic calculus starts are never for the whole society 'given' to a single mind which could work

⁵² Baier 1958, pp. 149-150.

⁵³ For our purposes, we need not require that moral rules be overriding in *every single case* but they must be overriding generally.

⁵⁴ This reason is of course not the sort of reason that would be offered as a justification for particular rules. Instead, I'm merely explaining why we need a set of moral rules from the perspective of philosophical theorizing.

out the implications and can never be so given.”⁵⁵ No one can collect all the relevant information about individuals’ plural wants, goals and reasons. The “moral economy” faces a similar problem: individuals need a set of publicly justified rules to make social life possible but no one can collect all of the information relevant to determining which norms are justified to all. These informational limitations illuminate the challenges of N-person public justification. A society that aims at public justification requires publicly justified norms despite the fact that no one will know *why* they are justified to many of those subject to them.⁵⁶

Moral rules help solve Hayekian worries. On Darwall’s view, it often appears that particular moral demands are the objects of justification. Thus, each time a moral demand is made it must be justified through the issuing of a second-personal reason. If moral justification is to occur through an exchange of reasons, then this micro-level process must be scaled up to the societal level. But moral justification cannot work this way. Instead, individual actions must be replaced by some mediator, or a social practice that obviates the need for explicit moral justification. In economic coordination, prices serve this function. Individuals need not engage in a vast array of dyadic bartering. Instead, they can rely on the information provided by prices to regulate their activities without knowing much about the factors reflected in those prices. Moral rules can play the same role in moral cooperation.

Darwall can give two replies. First, perhaps Darwall can help himself to justifications for certain *types* of moral demands such that reason-giving occurs in terms of types of moral demands rather than the demands themselves. Thus, if John catches Reba breaking the “don’t lie” type of demand, he needn’t justify each moral demand he makes but instead offer reminders to Reba that she knows better than to lie in general and thereby reference some previous act of intersubjective justification. Depending on how finely “types” are individuated, I might agree with Darwall here. If types are individuated thickly enough, then Darwall will simply have built an account of social-moral rules into his theory. Address will be mediated via types or rules with at least some (though not all) of the features that I have cited. The problem with Darwall’s theory will then become that the account of rules contained in his view is undeveloped. My criticism of Darwall then will simply be that further focus on moral rules will lead to important modifications of his view.

⁵⁵ Hayek 1945, p. 519.

⁵⁶ The thin account of publicity I give here is intended to fit with the convergence-moderate idealization account I defend in Part II. Convergence permits many types of reasons to interface with a social-moral rule and moderate idealization helps to show how the information problem can be solved.

However, Darwall might still be suspicious of rule-mediated moral address. He might argue that rules cannot mediate second-personal address because the dignity of persons requires that we address our fellows directly. For John to make a legitimate demand of Reba, he must address Reba by offering justifications to her. Perhaps John and Reba acknowledge common moral rules as binding and when John identifies Reba's infraction of a rule, he addresses her in a second-person way by making a demand of Reba that he believes she is committed to obeying. Thus, John respects Reba by making this demand of her but he still must address her directly. This reply is compatible with my account, though it still makes moral address too intimate. We frequently morally address others that we hardly know and cannot plausibly be argued to be within our moral community. The demand can be verified as legitimate independently of knowledge of the particular person. Social-moral rules are much like traffic laws: they allow us to get along even when we do not know one another. Traffic laws are publicly acknowledged rules and we frequently blame people for violating them even when we cannot see them through their (perhaps) tinted windows. I honk my horn at those who violate traffic laws and I have no idea who they are. This seems to be an appropriate form of moral address, one that is only appropriate in the context of commonly recognized moral rules, but it does not seem to involve the dyadic model of justification on which Darwall models the second-person standpoint.

Moral rules can arise through many processes, but they cannot always arise through explicit, deliberative cognition. As we have seen, all societies have some acknowledged set of moral rules. Accordingly, members of those societies will already regard many of these moral rules as binding. But if reason exchange is required for moral justification, then citizens must be prepared to offer reasons to others to justify new rules, revise old ones and punish violations with ostracism and blame to an implausibly great degree. This process substantially ties down the process of moral justification as it would the generation and enforcement of intersubjectively justified traffic laws. In this way, explicit, deliberative processes of moral justification are simply *too cumbersome* to generate the vast array of moral rules present in modern extended orders. Individuals rarely know the reasons of others and thus will only rarely know how to publicly justify their claims upon them. The knowledge of reasons is too dispersed.⁵⁷ For this reason, norms must be capable of being justified and even enforced without an exchange of reasons. Fortunately, the Public Justification Principles

⁵⁷ I take it that this problem remains even under conditions of hypothetical reason-giving. Any plausible hypothetical model of reason-giving will need to grapple with how hypothetical agents solve these information problems. Someone might reply that the idealization will eliminate the information problems but I argue against this radical form of idealization in Chapter 9.

holds that rules are justified whenever each person subject to the rule has conclusive reason to endorse it. It therefore says nothing about how moral rules must arise or how they must be justified. Consequently, deliberation and reason exchange are not essential to moral justification.

Moral rules permit members of extended moral orders to know *that* rules are justified without requiring that they know *why*. Moral rules “summarize” facts about whether a rule is publicly justified and thereby avoid demanding requirements for the public justification of social morality. The fact that moral rules are practiced typically indicates that a large range of people regard them as legitimate. While some people will obey a rule they regard as unjustified, it is unlikely that a moral rule could exist as a social fact if a sufficiently large number of people thought the rule was illegitimate.⁵⁸ We can compare the existence of moral rules with the existence of other norms, like norms of monetary usage. If no one believed that dollars were money in the United States, then dollars would cease to be money. And if everyone subject to a moral rule regarded it as a false rule, then it would cease to be practiced. In this way, the very presence of moral rules provides some evidence that moral rules are justified. They provide especially strong evidence if they are not disputed by any social group and are regularly followed without widespread violation. Citizens will never know why an extant and uncontroversial moral rule is justified to each person subject to it but they can acquire evidence that a large range of people regard the rule as justified. This is *not* to say that moral authority itself is conventional, rather that the mediators of second-personal authority (rules) are conventional.

To further illustrate, we might compare the summarizing function of moral rules to Rawls’s third stage of political justification, “public justification” (not to be confused with our use of the term). Rawlsian public justification occurs when “all the reasonable members of a political society carry out a justification of the shared political conception of justice by embedding it in their several reasonable comprehensive views.”⁵⁹ Rawls emphasizes that public justification only depends on reasonable comprehensive doctrines in an “indirect way” because “citizens do not look into the content of others’ doctrines.”⁶⁰ Instead, they are only concerned with “the existence ... of the reasonable overlapping consensus itself.”⁶¹ In Rawlsian public justification, individuals need not know why an overlapping consensus exists, so long as they know *that* it exists. Rawlsian public justification can be achieved if each person has found a place for the political conception within her

⁵⁸ Doubtless this claim is controversial. I discuss it in much more detail in Ch.10, II.v and III.iii.

⁵⁹ Rawls 2005, p. 387.

⁶⁰ Ibid.

⁶¹ Ibid.

comprehensive doctrine. Citizens only need to know that this process has occurred; they need not know the details. In my view, moral rules play the role of the political conception in public reason.⁶²

III.iv The Challenge of Publicity

Yet Rawls's conception of publicity presents a challenge to the mild form of publicity I defend. Let us first examine Rawls's complex ideal of publicity. On Rawls's view, publicity has three levels. The first level "is achieved when society is effectively regulated by public principles of justice: citizens accept and know that others likewise accept those principles, and this in turn is publicly recognized."⁶³ Everyone understands that a society's basic structure is justified in terms of commonly accepted practices of inquiry and reasoning. No problem is posed by this level of publicity since it only requires knowing that the principles of justice regulating society's basic structure are accepted and recognized by all. The second level of publicity, though, requires that citizens of a well-ordered society have shared beliefs "in the light of which first principles of justice themselves can be accepted, that is, the general beliefs about human nature and the way political social institutions generally work, and indeed all such beliefs relevant to political justice."⁶⁴ Depending on how general these shared beliefs are, the second level of publicity may not run afoul of Hayek's challenge. For instance, if publicity involves knowledge that the political conception of justice is accepted on the basis of the commonly acknowledged equality of persons, no problem arises. The third level of publicity concerns what Rawls calls "the full justification of the public conception of justice as it would be presented in its own terms" and must include all of the considerations we bring to the table when generating a conception of justice in the first place.⁶⁵ The full justification must be "publicly known, or better, at least to be publicly available."⁶⁶ Rawls allows for the weaker second condition because many may not want to expend the cognitive resources required to acquire the relevant knowledge. When a well-ordered society has all three forms of publicity, Rawls claims that it meets "the full publicity condition".

⁶² I cannot argue for this here but it is better to disaggregate a political conception of justice into an array of moral rules if for no other reason than to permit more complex public justifications in large, modern, pluralistic societies.

⁶³ Rawls 2005, p. 66.

⁶⁴ *Ibid.*, p. 67.

⁶⁵ *Ibid.*

⁶⁶ *Ibid.*

Full publicity is an important ideal for public justification (in our sense) because it is appropriate to our status as “reasonable and rational citizens who are free and equal.”⁶⁷ Full publicity is required to know whether a society’s basic structure is based on unjustified coercion. So in contrast to Rawls’s claim that we only need to know *that* an overlapping consensus exists, it appears that if citizens cannot know the reasons why their society’s basic structure is justified, it is hard to see how they can know *whether* it is justified. In contrast, if a political conception of justice possesses full publicity, the exchange of reasons in a free society will render citizens “confident that this avowed reckoning itself will strengthen and not weaken public understandings.”⁶⁸ For Rawls, it is vital that a just society not depend upon delusions and historical accidents. While Rawls denies that there can be any certainty about whether society’s basic structure is so based, full publicity still requires that “in [citizens’] public political life nothing need be hidden.”⁶⁹ If we answer Hayek’s challenge by abandoning full publicity, then we abandon knowing whether our society is just. It thereby appears that to adopt an N-person account of public justification through the emergence of moral rules requires abandoning the great good of full publicity.

A system of moral rules can still achieve full publicity, however. Seeing why requires distinguishing between two ways of understanding the process of public justification. One might imagine that public justification requires showing how each and every norm in a society is publicly justified to each individual. On this view, the public reason project involves reverse engineering all of a society’s norms in order to see whether and how they are justified. In contrast, one might conceive of the public justification project as taking our current practices for granted as *pro tanto* justified. Public reason liberals instead search for public justification in response to *complaints*. Public justification thus becomes a method of resolving real-world disputes. It is arguable that the early Rawls held the second view, given that he maintained that “perhaps the principle aim of ethics is the formulation of justified principles which may be used in cases wherein there are conflicting interests to determine which one of them should be given preference.”⁷⁰ If we think that full publicity can only be achieved by the former method, Hayek’s challenge becomes formidable. But if the public reason project is only taken to require resolving disputes as they arise, then full publicity can be achieved for norms in dispute. As a society evaluates its normative practices over time, full publicity can be gradually achieved for a society’s normative practices on the whole. Hayek’s challenge is

⁶⁷ Ibid.

⁶⁸ Ibid., p. 68.

⁶⁹ Ibid.

⁷⁰ Rawls 1951, p. 186-7.

thereby met since the latter method of justification requires far less effort to collect and evaluate information about citizens' reasons. Hayek decried the attempt to uncover the whole rationale behind the structure of a society's political and economic institutions due to the informational challenges such a reconstruction faced. Instead, Hayek conceived of political society as refining, extending and correcting the evolved moral, political and legal norms of that society as a whole. Hayek spent most of his career attacking what he called "constructivist rationalism"—"a conception which assumes that all social institutions are, and ought to be, the product of deliberate design."⁷¹ Rawls felt similarly, claiming that the social practices we are evaluating are "already established, there is no question ... as to how they will set up these practices for the first time."⁷²

Samuel Freeman thinks there are additional problems for limited publicity. Publicity allows that "moral agents can know the real reasons for moral constraints and expectations and can apply these reasons to plan their actions and pursuits" which is in turn required to be "a fully responsible moral agent."⁷³ With publicity, citizens can form stable expectations that their institutions will not turn on them because these institutions are rooted in a genuine public consensus. And yet, Freeman claims that with full publicity "even the best of people ... can be mistaken about the nature and requirements of moral reasons" and "accept different reasons in justification of their own and others' morally justifiable conduct."⁷⁴ These mistaken beliefs, "... are not a reason for denying humanity's capacity for justice and its moral nature. ... Nor should their mistaken beliefs alone lead to the instability of a well-ordered liberal society, so long as there is an overlapping consensus"⁷⁵ It appears that publicity can be achieved even if citizens have mistaken beliefs about why moral rules are justified so long as an overlapping consensus exists. Thus moral rules seem able to provide stable expectations even if we do not know the reasons they are justified. While providing for expectations is only one function of moral rules, it is a vital one.

We can conclude, then, that our account of moral rules can survive the challenge of publicity. And we can thereby draw a more general conclusion: that the mainstream version of contractualism is weakened by omitting an account of moral rules. Without an account of moral rules, contractualism cannot make sense of how public justification occurs across large populations

⁷¹ Hayek 1973, p. 5.

⁷² Rawls 2001, p. 53. Aaron James maintains that this was always Rawls's view. See James 2005. For further discussion of this point, see Ch.10, III.i.

⁷³ Freeman 2007, p. 6.

⁷⁴ *Ibid.*, p. 33.

⁷⁵ *Ibid.*

of individuals who cannot engage in dyadic justifications with one another. Thus one function of a society morality of rules is to solve N-person problems of moral justification.

III.v Moral Rules and Second-Person Reasons

Moral rules make sense of public justification across large groups. Accordingly, moral theories that understand public justification dyadically cannot capture the complexities involved in society-wide public justifications. This is a weakness in mainstream contractualism. However, Darwall's theory is subject to a further problem. Since his second-personal conception of moral justification lacks a well-defined place for moral rules, the justificatory force of second-person reasons becomes obscure. Scanlon's theory largely avoids this concern because Scanlonian moral justification involves "principles," though these are distinct from moral rules. Despite his dyadic account of public justification, Scanlon's account of principles might be expanded to accommodate the concern I press against Darwall below.

Second-personal reasons are those essentially concerned with how one should act with respect to others. Darwall seems to imagine public justification occurring by reasons somehow "reaching across" persons, that is, somehow moving from one person to another. For Darwall, when we bind the wills of others, we *issue* second-personal reasons that they regard as having force. But what it means to issue such a reason is obscure. We can *utter* a reason in the form of a proposition and remind others of their reasons, but how can second-person reasons obligate merely by being uttered? Darwall recognizes that he must explain what it is about a moral demand that makes it genuinely moral. His claim is that bona fide moral demands are those made in accord with reasons with second-personal content. But it is unclear what it is for a reason to have second-personal content. Perhaps the "second-personality" of a reason consists in it referencing another person. Or perhaps a reason's second-personality consists in its being somehow sensitive to the reasons of others.

Darwall could argue that the domain of second-person reasons is marked off by their appropriate use within accountability relations. In other words, what makes a second-person reason second-personal is the context in which it is employed. This interpretation of Darwall's view seems to belie the text, as Darwall focuses on explaining what the content of a second-personal reason consists in (like its agent-relativity). Yet even on this more contextualist view of what makes a reason second-personal, it seems like Darwall has not said enough. Reasons can be offered within the

context of accountability relations all the time, but that does not seem sufficient to make them second-personal. I can offer, for instance, reasons of the wrong kind to others within accountability relations; they just fail to bind the wills of those to whom they are addressed. And so the question remains: what is it about the *content* of the reason that causes it to oblige the conduct and will of others?

To my mind, Darwall mistakenly builds the “other-regarding” second-personal aspect of a public reason into the *content of the reason* rather than *its place in a rational being’s epistemic and moral commitments*. To illustrate the distinction, recall how moral rules mediate public justification. A rule is proposed as an object of justification. It is publicly justified when each person has a conclusive reason of her own to endorse the rule. But any reason to which those subject to the rule are committed will suffice. In fact, each person can have distinct reasons to endorse the rule. In this case, the *reasons* aren’t second-personal in Darwall’s sense. Rather, the *rule* is second-personal because it applies to all members of the moral community and is the basis upon which we address others. We can make sense of Darwall’s claims about the nature of the second-person standpoint if it is the standpoint we take when we propose and enforce publicly justified moral rules. On this view, second-personal address occurs when John demands that Reba conform her behavior to a rule he knows is justified to her. John respects Reba’s dignity—he treats her as free and equal—when he recognizes that his moral demands of her must be made in accord with a rule that is publicly justified to her. John blames Reba appropriately when she fails to behave in accord with publicly justified moral rules. Darwall has made a valuable contribution to the public reason tradition in showing that an attractive account of right and obligation can be derived from the public justification of social morality and that the process of public justification rests upon social practices that we all regard as binding. But the notion of a second-person *reason* need not figure in a theory of public justification.⁷⁶

Moral rules explain where to locate the normative force of a public reason. Without an account of moral rules, it is unclear how second-person reasons can be justificatory. Accordingly, Darwall’s account of the second-person standpoint tries to explain the second-personality of reasons by identifying second-personality with some property of individual reasons. In Chapter 4, I argue that a better account of the second-personality of reasons can be developed by identifying the

⁷⁶ Were Darwall to reply that second-person reasons are only second-personal when issued in a second-personal context of accountability relations, we would as of yet not have a sufficient condition for a reason being second-personal. We can live in accountability relations with others and issue all kinds of reasons, ones that Darwall would likely not count as second-personal.

appropriate relation those reasons bear to a rational being's epistemic commitments. Specifically, I will argue that a reason counts as second-personal when it is either a member of one's belief-value set or can be soundly inferred from one of those members and when it concerns the practice of social morality. Second-personal reasons are reasons that require obeying and disobeying moral rules. The account developed in Chapter 4 will therefore abandon the search for a reason-content that specifies whether a reason is second-personal, thus altering Darwall's account of public justification.

IV: Modifying a Contractualist Account of Liberal Practical Reason

Mainstream contractualism has two flaws: (i) it packs too much substantive content into the idea of demanding treatment, and (ii) it lacks an account of moral rules, causing it to locate the second-personality of reasons in their contents rather than their relations to other reasons and rules. Given these two flaws, mainstream contractualism cannot provide an attractive model of liberal practical reason and therefore cannot form the structure of a strategy for reconciling public reason liberals and their integralist critics.

These errors must be rectified. In Chapter 4, I develop a model of liberal practical reason that roots public justification in a presumption against interference and a relational and rule-based account of second-personal reasons. We will see that this modified version of contractualism provides an attractive model of liberal practical reason that explains how public justification and integrity relate to one another. A promising strategy for reconciliation will emerge. For instance, it will become clear that the accessibility and shareability requirements in public reason liberalism are motivated in part by the same confusions that led to locating the second-personality of reasons in their content. In Chapter 6, I will show that when public reason liberals try to explain how reasons can justify coercion, they look to the content of the public reason for its public normative force. I suggest an alternative model of public reasons in Chapter 7 which understands the publicity of reasons in terms of their connection to publicly recognized rules and proposals. Locating the second-personality of reasons in their content, I will argue, motivates both the consensus conception of public justification and what I call the symmetry requirement—two aspects of the mainstream theory of public justification that make resolving the role of religion debate seem intractable.

Chapter 4: A New Model of Liberal Practical Reason

Chapter 3 showed that mainstream contractualism cannot provide a satisfying model of the liberal dualism of practical reason that characterizes the conflict between public reason liberals and integralists. But the failures outlined in Chapter 3 are instructive. Chapter 4 claims that by revising the mainstream contractualist model of liberal practical reason, we can construct a strategy for resolving the religion and politics conflict. Consequently, this chapter must proceed in three general stages: (i) specify an alternative connection between respect for persons and public justification, (ii) develop an account of first and second-person reasons that depend on an account of moral rules, and (iii) show how the resulting structure of practical reason provides a strategy for resolving the religion and politics conflict. The dissertation stands or falls on Chapter 4. Without Chapter 4, the strategy employed throughout Part II will fail. But if Chapter 4 is a success, then we can enter Part II with a strategy for resolving an old and powerful problem.

Chapter 4 meets its three goals in six parts. Section I outlines the idea of a presumption against interference and demonstrates its superiority to the standard of demanding treatment. Section II explains what it means for a reason to be public. Section III articulates a theory of second-personal reasons as public reasons concerned with following, rejecting or violating moral rules. Section IV develops a theory of first-personal reasons and connects them with integrity. By the end of Section IV, we will have the resources for representing the relationship between the Public Justification Principle and integrity-based reasons. Section V characterizes the relationship between first and second-personal practical public reasons and briefly sketches the strategy for reconciling them. Section VI concludes by sketching the strategy in the Introduction to Part II.

I: The Presumption against Interference

The standard of demanding treatment that lies at the heart of contractualist moral theory should be replaced by a *presumption against interference*. The presumption against interference holds that each person has a *prime facie* liberty-right to act on her reasons as she understands them and that those who wish to interfere with her must justify their interference by meeting the presumption. The presumption against interference mirrors liberal political theory's presumption in favor of liberty and

thus can represent the presumption in favor of liberty as a subset of a presumption within practical reason itself. Defending the presumption against interference requires (a) explicating the notions of interference and a presumption, (b) motivating the presumption and (c) showing how the presumption against interference resolves what the standard of demanding treatment cannot.

I.i Interference

Interference comes between a person and the actions she wishes to perform by interrupting her *self-direction*. Stanley Benn and Gerald Gaus have argued that the notion of a self-directed person forms the foundation of the liberal theory of the person.¹ Consequently, constructing a theory of liberal practical reason should start with the minimal assumption that human beings are self-directing or *autarchic*. For Benn, an individual is autarchic when she,

- (i) possesses the powers and capacities that enable her to act;²
- (ii) faces opportunity costs when choosing a course of action;
- (iii) has beliefs about these capacities and costs; and,
- (iv) has aims regarding states of affairs she would like to achieve that she regards as worth the cost.³

According to Gaus, the vast majority of modern humans consider themselves self-directing in Benn's sense. They conceive "of agency as an expression of the self ..."⁴ Persons are in this way "individualized sel[ves]."⁵ To deny this, we must engage in "a sort of self-destructive act" since this conception of the self is fundamental to our identity.⁶ I agree. Self-direction is a clear enough notion and common to a large enough group of human beings that it is a safe assumption upon which the build a theory of practical reason.

¹ Gaus has argued that the entire modern liberal theory of man develops out of an attempt to "[combine] and [reconcile] individuality and sociability." Gaus 1983, p. 2.

² Without the room to cash this out in detail, let us say that these powers and capacities include physical powers to act with one's body upon some object, the ability to make rational choices on a fairly regular basis and the ability to translate decisions into action. A rational person usually lacks defects in epistemic rationality, such as when one has been brainwashed, or when one is practically irrational due to some neurosis, or even when they have psychic discontinuity or suffer from psychopathy. For most discussion see, Benn 1988, pp. 156-164.

³ Ibid., pp. 152-3.

⁴ Gaus 1990, p. 385.

⁵ Ibid., p. 384.

⁶ Ibid., p. 383.

The concept of interference employed in this dissertation holds that to interfere with John is to reduce John's set of choice options *qua* self-directing being. All else equal, offers are not cases of interference as they increase John's option set, his opportunities to direct himself. Should John have a range of options to choose from and Reba reduces this range with the intention of changing his behavior, then she has interfered with him.⁷ Further, our notion of interference is non-moralized and so does not by itself imply that interference is good or bad, justified or unjustified. Public reason liberalism and contractualism both propose to explain how many moral and political concepts should be understood. To use a moralized conception of interference would thus load too much theoretical content into the foundation of the view. This would be to repeat the error of the standard of demanding treatment discussed in Chapter 3.

Actions and moral demands that reduce option sets interfere with agents.⁸ For actions this is clear. If John is building a house and Reba takes his materials, she has interfered with him. But how can a claim or demand interfere? If John makes a demand of Reba, she can ignore him. Yet moral claims and demands constitute interference for two related reasons. First, normally functioning human beings already acknowledge moral demands as having force, or as Baier would say, moral rules are enforced with a "characteristic moral pressure".⁹ According to Darwall, human beings implicitly acknowledge reciprocity of claim. We respond to the demands of others because we recognize their dignity as rational agents. It is our status as members of the human moral community that leads us to recognize the normative force of the claims we make on one another's conduct and will. Normally functioning agents, therefore, care about being moral and care about their relations with others. Violating moral demands puts us at odds with our communities. Second, since moral demands threaten to (at least partly) strongly sever our ties to the moral community, the cost of ignoring moral demands can limit our options to act in ways that violate widely acknowledged moral norms. Thus, when we think claims upon us justified, the claim renders an option morally illegitimate. As Gaus puts it, "a moral rule against stealing means that Alf has a duty not to steal, so

⁷ As with self-direction, my conception of interference largely follows that of Stanley Benn. See Benn 1988, pp. 122-151. Note that while cases of interference are vast in any complex society, the number of moral rules that justify it are similarly vast. As we saw, no act of explicit justification must occur for every case of interference. So much of the justification of interference is implicit. I thank Christopher Eberle for pressing me on this point.

⁸ Note that a moral demand is not the same as making a moral judgment. If Reba says to John, "What you have done is wrong", this is not by itself an act of interference. Instead, Reba interferes when she engages in the speech act of blaming, when she says, "Hey, John, stop that." Or "Hey, John, you're being a jerk." Thus, Reba interferes when she issues a prescription, not merely when she makes a mere declarative statement. I thank Nicholas Wolterstorff for pressing me on this point.

⁹ Baier 1954, p. 108.

stealing is not a legitimate option for Alf: he is not at liberty to steal.”¹⁰ An otherwise attractive option might be closed off by a justified moral rule.

That said, moral claims and demands can interfere even when they are not justified. Human agents worry about being sanctioned and so often worry about the threat of enforcement or punishment. Violating accepted moral rules usually results in the often unpleasant and frequently devastating experience of being *blamed*. In this way, our moral practices contain implicit threats of punishment or sanction, as we have seen. When John threatens Reba with blame he thereby reduces her option set even if Reba commits no moral infraction.¹¹ As J.S. Mill famously noted, “Society can and does execute its own mandates; and if it issues wrong mandates instead of right ... it practices a social tyranny more formidable than many kinds of political oppression.”¹² Social enforcement of morality is like state coercion. Both decrease options, interfere and have the capacity to, at their worst, ruin human lives.

The reader will notice that I have built the dissertation’s conception of interference upon the idea of interrupting self-direction or autarchy rather than a more substantive notion of autonomy. I do so for a reason. The point of developing a model of liberal practical reason is to develop a conception of public reason liberalism that can answer the integrity objection. Above (Ch.2, V) I claimed that the only way to do so is to find a form of public reason liberalism that can show how the restrictions placed upon citizens of faith by public reason liberalism are self-legislated. Chapter 4 attempts to model what self-legislation or rational autonomy consists in. Citizens are self-legislating and autonomous when all coercion and interference is publicly justified to them. But Chapter 4 advances a model of public justification. Thus to root part of the theory of liberal practical reason (the presumption against interference) in a substantive notion of autonomy would get things the wrong way around. Autarchy and interference are concepts out of which a conception of rational autonomy should be constructed.

In sum, John interferes with Reba when his actions or demands reduce her option set vis-à-vis her self-direction. Fortunately we need not develop the concept of interference employed in the dissertation further since the forms of interference relevant to the religion and politics conflict count as interference on most moral and political views. Let’s turn to the idea of a presumption.

¹⁰ Gaus 1999, p. 5.

¹¹ For an explanation of this form of negative liberty as an “opportunity concept”, see Gaus 2000, pp. 81-82.

¹² Mill 1978, p. 4.

I.ii Presumptions

A presumption within moral theory marks out justified norms from unjustified norms; it identifies a set of justified moral rules. Social morality is full of presumptions; one prominent example is the presumption of innocence within trial law. The presumption of innocence bears a crucial similarity to the presumption in favor of liberty. The presumption of innocence does not provide a reason to suppose that the defendant in a court case is innocent. Instead, it structures judicial practice. It *specifies how a successful and justified trial must proceed*. Unless the prosecutor can show that the defendant is guilty beyond a reasonable doubt, she fails to successfully prosecute a crime. As explained above (Ch.1, II.i), public reason liberals give the presumption in favor of liberty a special status. The presumption in favor of liberty, like the presumption of innocence, structures the practice of the justification of coercion. When the presumption is met, coercion is justified and otherwise not. The presumption in favor of liberty and its associated moral equivalent, the presumption against interference, are thereby distinct from other presumptions within political and moral theory. They are not mere parts of their respective field of norms. Instead, both presumptions must be met *to generate* norms—to make them bona fide. In other words, only when a reason meets either presumption can the claim the reason supports be justified. A presumption, again, marks out the set of justified moral norms by evaluating whether each norm can meet the presumption. Justified norms meet the presumption; unjustified norms do not.

I.iii The Presumption Against Interference

The presumption against interference is built on the assumption that practical reason contains an *asymmetry* between the permissibility of self-directed action and the permissibility of interfering with the autarchic action of others. Self-directed action possesses a default position within morality; it thereby implies that a natural liberty-right to action must be rebutted. Persons are permitted to act on their own reasons unless they have some recognizable reason to the contrary. This is to say that persons need not justify acting on their reasons unless they interfere with others in doing so. Benn illustrates the asymmetry of moral justification as follows:

Imagine Alan sitting on a public beach, a pebble in each hand, splitting one pebble by striking it with another. Betty, a casual observer, asks him what he is doing. She can see, of

course, that he is splitting pebbles; what she is asking him to do is to explain it, to redescribe it as an activity with an intelligible point, something he could have a reason for doing. There is nothing untoward about her question, but Alan is not bound to answer it unless he likes. Suppose, however, that Betty had asked Alan to justify what he was doing or to give an excuse for doing it. Unlike explanations, justifications and excuses presume at least *prima facie* fault, a charge to be rebutted, and what can be wrong with splitting pebbles on a public beach? Besides, so far as we can tell, Alan is not obliged to account to Betty for his actions

Suppose Betty were to prevent Alan from splitting pebbles by handcuffing him or removing all the pebbles within reach. Alan could now quite properly demand a justification from Betty, and a *tu quoque* reply from her that he, on his side, had not offered her a justification for splitting pebbles, would not meet the case, for Alan's pebble splitting had done nothing to interfere with Betty's actions. The burden of justification falls on the interferer, not on the person interfered with. So while Alan might properly resent Betty's interference, Betty has no ground for complaint against Alan.¹³

If Betty wishes to interfere with Alan, Alan is owed a justification. It should be stressed that in Benn's case, Betty is interfering with Alan through her actions not her claims. She simply prevents Alan from splitting pebbles by collecting the pebbles or handcuffing Alan. But Betty could just as well interfere with Alan by demanding that he stop splitting pebbles, either through asserting her authority over him without reason or by giving him a reason. Gaus sometimes understands the presumption against interference as a "right to natural liberty" which he claims "articulates the basic moral assumption of self-directed agents, an assumption so basic that it can only be denied by a sort of self-destructive act."¹⁴ To justify interference to the one interfered with is to recognize this fundamental mode of conceiving of the self that is shared (at least) across all Western societies. In Scanlon-Darwall terms, we thereby recognize each other's dignity as practically reasoning beings, as persons guided by their deliberations, or as Benn puts it, persons with "inner-impelled personalities."¹⁵

¹³ Benn 1988, p. 87.

¹⁴ Gaus 1990, p. 383. This view is not confined to modern liberals, however. Locke claims that "all men are naturally in ... a state of perfect freedom to order their actions ... without asking leave, or depending upon the will of any other man." Locke 2003, p. 101.

¹⁵ Benn 1988, p. 162.

It is crucial that we understand the presumption against interference as a presumption that must be met in order to show that a moral demand or interfering action is justified. The presumption against interference serves as a “rite of passage” for proposed moral norms. When a proposed moral norm overcomes the presumption against interference, like Pinocchio, it becomes “a real boy”. Some will object to privileging the presumption against interference over presumptions in favor of other values, such as a presumption against causing harm or a presumption in favor of helping those in need. But if we understand the presumption against interference as a rite of passage we avoid this objection, since the different weights of values are set within social morality *as regulated by* the presumption against interference. On analogy with the legal presumption in favor of innocence, we do not employ the presumption in favor of innocence as evidence. Instead it structures our practice of weighing evidence. We don’t need the presumption in favor of innocence to be proven by means of judicial practice. It is a proper part of judicial practice.

This manner of conceiving of the presumption against interference raises a complication. If the presumption against interference is “outside” the realm of social rules, guarding the gate, then how can it be a norm at all? And if it is “inside” this realm, then how can liberals assign it such privilege over other presumptions? To answer these questions, we must distinguish the sense in which the presumption against interference is both inside and outside of social morality. The presumption against interference is outside of social morality since it is an expression of the concept of the person that lies at the *ground* of social morality, specifically the idea of an autarchic or self-directed individual. But the presumption against interference can derive its normative status from being justified to members of modern Western societies in accord with their shared ideas, in particular their shared idea of the individual as self-directing. In this way, defenses of the presumption against interference like Gaus’s can be understood as a public justification for the presumption.¹⁶ While the presumption against interference is grounded in the structure of practical reasoning among self-directing individuals, recognizing this fact does not show that the presumption against interference need not be justified. Instead, we can understand the argument that the presumption against interference is grounded in the structure of practical reason *just is* the public justification to employ it. The argument for the presumption in favor of interference is itself the rationale or reason that satisfies the Public Justification Principle.

¹⁶ Gaus 1990, pp. 381-90. Rawls actually has a similar view. See Rawls 2005, p. 100.

Some find the presumption trivial. For instance, Joseph Raz argues that a simple presumption in favor of liberty is “very weak” and can be endorsed by non-liberals.¹⁷ Such a weak presumption is thought to be unable to weigh different liberties according to their relative value. But if we understand the presumption against interference as a gatekeeper, we need not be concerned with this objection. The presumption against interference is only a *necessary condition* for the justification of social morality. The sufficient condition for the exercise of interference is the Public Justification Principle, which shows how the presumption against interference is met. Once considerations of reasons relevant to public justification are brought to bear on the problem of weighing liberties, a vast array of rankings of liberties becomes possible.

I.iv The Presumption against Interference vs. Demanding Treatment

Recall (Ch.3, II.i) that Darwall and Scanlon base their respective contractualisms on the vague notion of demanding treatment. Again, neither Scanlon nor Darwall provide an argument in favor of the move from human dignity to the notion of demanding treatment and justifiability to others. Not only can a substantive conceptual gap be driven between dignity and moral justification (Ch.3, II) but the attempt to do so reveals immediate and deep complexities. In contrast, the presumption against interference is a far clearer idea. It appeals to philosophical concepts that despite needing analysis (which moral and political concepts don't?) at least provide a plausible path to careful analysis. We as of yet have no reason to think that the presumption against interference cannot be employed as part of a theory of liberal practical reason. So, one compelling reason to adopt the presumption against interference over the standard of demanding treatment is its clarity.

Second, the presumption against interference is more clearly rooted in the moral personality of human beings. The idea of human dignity is fleshed out in terms of respecting the conception of ourselves that human beings almost universally share—our conception of ourselves as self-directed. The way that dignity is respected—the way that we respect *persons*—is by treating them as subjects of justification by meeting the presumption against interference, thereby expressing respect for our common conception of ourselves as self-directed beings. Benn claims that this form of recognition lies at the heart of practical reason, that when we look to the nature of practical reasoning itself, we

¹⁷ Raz 1986, pp. 11-14.

will see the asymmetry of moral justification upon which the presumption against interference is based.¹⁸

The presumption against interference can serve as a moral foundation for the presumption in favor of liberty within public reason liberalism. Recall (Ch.1, II.i) that the presumption of liberty is well-grounded in the history of liberal political theory. Locke speaks as a representative of the classical social contract theorists when he claims that “all men are naturally in ... a state of perfect freedom to order their actions ... without asking leave, or depending upon the will of any other man.”¹⁹ As we have seen (Ch.1, II.i), the presumption in favor of liberty is endorsed by modern liberals like Feinberg, Rawls, Benn and Gaus. The presumption against interference has an obvious connection to the presumption in favor of liberty. The presumption in favor of liberty is a presumption against political coercion, and coercion is a type of interference. While moral demands always interfere, they are not necessarily coercive. Instead, coercive actions and demands are a subset of those actions and moral demands that interfere. Thus, if practical reason contains a presumption against interference and if political morality is a part of social morality generally, then the presumption in favor of liberty is political theory’s corollary to practical reason’s presumption against interference. We can conclude, therefore, that the presumption against interference is clearer than the standard of demanding treatment and is more appropriate to public reason liberalism. In this way, it more clearly connects respect for persons and human dignity with public justification. These reasons appear sufficient to justify replacing the standard of demanding treatment with the presumption against interference.

II: A Theory of Justificatory and Public Reasons

Section II offers a theory of the atomic units of moral justification: justificatory reasons. The presumption against interference only establishes the form of moral justification, the way in which reasons justify norms. To complete a model of liberal practical reason, we need a conception of reasons as well. Section II develops an account of *justificatory* reasons, the reasons that can be utilized by a contractualist theory of moral justification and an account of *public* reasons.²⁰ By the end of

¹⁸ For further attempts to ground the presumption against interference, Gaus 1996, pp. 120-129 and Gaus 2011, pp. 202-258.

¹⁹ Locke 2003, p. 101.

²⁰ It must be emphasized that the set of reasons relevant to contractualism does not exhaust the set of normative reasons broadly speaking. Appendix A shows how this theory of public reasons cuts across a number of metaethical and epistemological distinctions that one might employ to divide normative reasons into classes (specifically the agent-

Section II, we will have a model of justificatory and public reasons upon which the crucial distinction between first and second-person reasons can be overlaid.

II.i Open Justification

Scanlon famously defined a reason for something as “a consideration that counts in favor of it” and thus takes the notion of a reason as a primitive.²¹ I shall take the notion of a reason as a primitive as well. Section II delineates the class of reasons relevant to the practice of social morality—justificatory reasons—and so something must be said about what makes justificatory reasons distinctive. The distinctive feature of a justificatory reason is that it can be “recognized” by those that “have” them. We must now explain precisely what this means. I shall understand this “rational recognition condition” in terms of the epistemic notion of *open justification*.

The concept of open justification can be understood in terms of the idea of a belief-value set developed above (Ch.1, IV.ii). Since public reason liberalism employs the notion of a belief-value set already, employing it to complete a model of liberal practical reason seems appropriate. John’s belief-value set is the set of all of his beliefs, desires, goals, and plans, i.e., everything he thinks and wants.²² We might call the elements contained in an agent’s belief-value set “mental primitives” or the fundamental elements of conscious and subconscious thought. On my view, all reasoning consists in the manipulation of these mental primitives. *Inferences* consist in attempts to derive new mental primitives from other ones. For instance, if John is considering whether he should believe in God, he might draw inferences from his beliefs about the order of the universe and his understanding of how the universe might have come to exist. If John is considering whether to take a lunch break, he will consult his desire for food and its weight relative to his desire to, say, continue working. He will then draw inferences from his beliefs about the relevant plan of action he must take to eat lunch and his beliefs about what a sound trade-off of desire satisfaction might consist in. If our mental primitives are the basic objects of reasoning, inference is the mental process of connecting these primitives in a sound manner.

neutral/agent-relative reasons distinction, the reasons internalism/externalism distinction and the distinction between externalist and internalist conceptions of epistemic justification).

²¹ Scanlon 1998, p. 17.

²² Again, the idea of a belief-value set derives from Bernard Williams’s conception of a subjective motivational set. See Williams 1981, p. 102.

A theory of justificatory reasons requires a method of sound inference, of how to rightly arrive at reasons for action or beliefs from our mental primitives.²³ I take a theory of sound inference to be a theory of epistemic justification. However, the term “justification” already has several uses in this dissertation, so I will understand epistemic justification as a form of *entitlement*.²⁴ I understand epistemic entitlement in terms of the *liberty to believe*. Thus, John is entitled to believe P if he has a liberty-right to believe P. In other words, he has a mere permission to believe P. In believing P, *John does no wrong*.²⁵ Similarly, John has a sound reason to X when he is entitled to affirm the reason. In this way, in affirming his reason to X, John does no wrong.

A conception of reasons for a theory of liberal practical reason must be *internalist*.²⁶ Contemporary epistemology contains numerous forms of internalist belief entitlement. One form of belief entitlement internalism is *access* internalism. Access internalism holds that whenever one is entitled to believe P one can become aware by reflection of all of one’s justifiers that P.²⁷ Externalist theories of belief entitlement deny access internalism (though they typically involve more than this). One such theory is Alvin Plantinga’s proper functionalist account of warrant, which holds that “a belief has warrant only if it is produced by cognitive faculties that are functioning properly in an appropriate environment.”²⁸ Hybrid views exist as well, such as William Alston’s internalist externalism, which holds that one is entitled to believe P when the belief is “based on an adequate ground.”²⁹ It is internalist in requiring that beliefs have a ground in order to be epistemically

²³ Or, if we are reasons externalists, we need a method of arriving at *beliefs* about our reasons for actions and belief.

²⁴ My notion of belief entitlement is influenced by the theory of entitlement employed by Nicholas Wolterstorff. See Wolterstorff 2009, pp. 86-117.

²⁵ In some cases, John is not merely permitted to believe P but required to believe P. But I shall understand being required to believe P in terms of permissions to believe P and alternatives to P. For instance, John is required to believe P when he is entitled to believe P and not entitled to believe any alternative to P (including merely \sim P and a suspension of belief about P one way or another).

²⁶ Or internalist with respect to justificatory reasons, not necessarily all reasons. In Appendix A, I argue not only that one can be an internalist about justificatory reasons without embracing an internalist conception of epistemic justification of moral reasons generally. I further argue that even reasons relevant to public justification might be all things considered epistemically justified in accord with a conception of epistemic justification that differs from the view I defend. The conception of epistemic justification I employ here is only relevant to the public justification of interference and coercion.

²⁷ Roderick Chisholm puts the view as follows: ‘We presuppose, second, that the things we know are justified for us in the following sense: we can know what it is, on any occasion, that constitutes our grounds, or reason, or evidence for thinking we know’. See Chisholm 1989, p. 17. Also see Pappas 2005 for other forms of justification internalism.

²⁸ Goldman 2008. Also see Plantinga 1993, p. 19. Warrant, for Plantinga, is the property of true beliefs that makes them knowledge. Thus, warrant is not identical with justification but plays a similar role. Plantinga defines his own view as follows: “a belief B has warrant for S if and only if the relevant segments (the segments involved in the production of B) are functioning properly in a cognitive environment sufficiently similar to that for which S’s faculties are designed; and the modules of the design plan governing the production of B are (1) aimed at truth, and (2) such that there is a high objective probability that a belief formed in accordance with those modules (in that sort of cognitive environment) is true.”

²⁹ See Alston 1988, p. 265.

justified.³⁰ Thus Alston also affirms a modest form of access internalism.³¹ However, Alston's view is externalist in that it construes the adequacy of a belief's ground in external terms: one need not be entitled to believe that one's belief has an adequate ground. Instead, the ground must *be* adequate.³² Thus, one can be entitled to affirm beliefs within one's belief value set according to access internalism, internalist externalism or proper functionalism. On access internalism, beliefs acquire entitlement if and only if the believer can become aware by reflection of all one's reasons for belief. On internalist externalism, beliefs achieve entitlement if and only if they are based on adequate grounds. And on the proper functionalist account, beliefs achieve entitlement when they are warranted and they are warranted if and only if they were produced by cognitive faculties that were functioning properly in an appropriate environment.

It is my view that a theory of liberal practical reason requires an access internalist conception of belief and reason entitlement.³³ First, if public reason liberalism is rooted in an even partly externalist conception of entitlement, it abandons the conceptual connection between internalist entitlement (what we might call "rational commitment") and what counts as a justificatory reason. Any externalist theory will characterize entitlement partly in terms of considerations that cannot be arrived at through sound inference. Consequently, such theories will imply that agents have reasons to which they have no rational access. This is to reject the rational recognition requirement on reasons. There is a second reason for this as well. Gaus argues that public reason liberalism requires a conception of epistemic entitlement that "shows how reasonable pluralism is possible."³⁴ In other words, reasonable and rational individuals must be able to have differing yet epistemically justified beliefs, some of which concern matters of great import. Gaus objects to externalist views on these grounds. One serious problem for an externalist theory like reliabilism is that "it supposes the notions of truth or correctness to be logically prior to justification."³⁵ Externalist theories make the entitlement of beliefs reliant upon their truth. Consequently, they render obscure how individuals

³⁰ Alston: "The view is internalist most basically, and most minimally, by virtue of the requirement that there be a ground of the belief." Ibid., p. 270.

³¹ Ibid., p. 272.

³² Ibid., p. 276. The adequacy of a belief's grounding depends "on its being a sufficiently strong indication of the truth of the belief grounded." Ibid., p. 270.

³³ It is important to emphasize again that access internalism is only the preferred conception of epistemic justification *for the public reasons relevant to the justification of interference and coercion*.

³⁴ Gaus 1996, p. 11.

³⁵ Ibid., p. 26.

can be entitled to have many false beliefs about morality.³⁶ A political theory that acknowledges reasonable pluralism must have a conception of epistemic entitlement that aids,

... in determining when we hold true beliefs and when we go wrong ... epistemology is in a deep sense ultimately practical: It provides guidance about how to seek truth and how to know it when we come across it. But all this supposes that we can give an analysis of justified belief that does not presuppose access to truth.³⁷

Remember Locke's point: the fact that a religion is true does not mean its adherents have the right to rule. And more generally: the claim that a moral proposition is true does not establish the authority of those who advocate it over those who reasonably disagree. Liberal practical reason requires that we acknowledge the differing values of others and respect them as independent reasoners with their own perspectives *regardless* of whether their views are true. For this reason, we must be prepared to acknowledge that others are entitled to affirm their ideals and values even if we consider them to be deeply confused. To be truly liberal, we must extend respect for persons even to those we think get the most important matters in life wrong. We must concede that they can be free and equal and *rational* beings. An externalist theory of entitlement suggests that citizens' reasons for action depend on whether they have true beliefs. In this way, externalisms do not help us justify moral authority and interference. They presume that (at most) someone already has the truth when the problem of moral justification is how to engage others when they *disagree about the truth*. The same restriction arguably holds against internalist externalism since it utilizes the idea of an adequate ground defined in externalist terms. A belief may have an adequate ground, but adequacy is still understood in terms of the probability that the grounds yield true judgments. Even this view ties epistemic entitlement too tightly to the truth.

A theory of liberal practical reason likely requires the version of access internalism Gaus terms "open justification." On Gaus's view, open justification "treat[s] [one's system of beliefs and reasons] as open to new information and arguments and, from this external perspective, mak[ing] judgments about what would then be justified in [one's system of beliefs and reasons]."³⁸ The key

³⁶ At least beyond the set of false beliefs that can achieve reliabilist justification through being the outputs of reliable but not perfectly reliable processes.

³⁷ Gaus 1996

³⁸ *Ibid.*, p. 31.

feature of open justification is that it “takes ALP’s current belief system as the point of departure.”³⁹ For Gaus, beliefs and reasons can acquire entitlement inferentially or foundationally and foundational beliefs can be generated by feelings or intuitions which give them *prima facie*, though weak, credibility.⁴⁰ Further, achieving entitlement for a belief “require[s] more than merely having a good supporting reason in one’s ... system; it requires holding the belief because of that reason.”⁴¹ Gaus thereby defends “a causal interpretation of ‘the basing relation’ Good reasons are relative to a system of reasons and beliefs.”⁴² All beliefs relevant to public justification can be accessed either as foundational beliefs or by means of inferential chains from those foundational beliefs. That said, open justification can be ascribed to beliefs based on information an agent may not presently possess but could possess under the proper circumstances. Liberal practical reason cannot settle for mere “closed justification” that restricts entitlement to “one’s present commitments....”⁴³ Individuals have rational commitments to beliefs and reasons that are not currently a part of their belief-value sets. To openly justify reason X to John is to show that it can be properly inferred from John’s present belief-value set conjoined with adequate information, reasoning ability and some degree of belief coherence.⁴⁴ Going with open justification permits us to engage in the process of idealization characteristic of public reason liberalism because there is a place within each citizen’s belief-value set where we can ascribe to them new information and better chains of inference.

To make the notion of open justification more concrete, let us apply it to three cases to show how three different types of beliefs and reasons acquire entitlement: foundational beliefs, inferred beliefs and *inferable* beliefs or beliefs that can be soundly inferred from one’s present belief-value set:

Case A (John Loves Dinner): It’s past quitting time for John. He feels hungry and rushes home to make sure his dinner is still warm. Given that John is often late home from work and has had to warm-up his dinner before, John’s hunger helps ground his belief that he has a reason to rush home. Accordingly, John’s reason to rush home is openly justified.

³⁹ Ibid., p. 32.

⁴⁰ Ibid., pp. 64-65, 83-84, 108-109.

⁴¹ Ibid., p. 63.

⁴² Ibid.

⁴³ Ibid., p. 35.

⁴⁴ Gaus describes his theory as “weakly externalist”, but he is not using the term “externalist” in the same sense that I have. See *ibid.*, p. 32.

Case B (John Loves School): John is a long-time school administrator who believes his school is underfunded. For this reason, John supports a property tax hike to finance an improved education system. John's experiences as a school administrator and his reasoning about how to improve public schools ground his belief that he has a reason to vote for the property tax hike. Accordingly, John's reason to vote for the tax hike is openly justified.

Case C (John Loves the Bomb): Despite working at a nuclear research facility, John opposes the use of nuclear weapons. While at work, he encounters a red button that he does not know will launch a nuclear missile. The relevant information concerning the button lies in a safety handbook that John ignored. The fact that the red button launches the missile can be easily inferred from his present belief-value set and a cursory perusal of the handbook. Were John to have the relevant information, he would take himself to have a reason not to push the red button. Therefore, John's reason not to push the red button is openly justified.

In each case, John has openly justified beliefs. Case A illustrates an openly justified reason whose entitlement derives directly from a perception of hunger and a modest amount of background knowledge. Case B illustrates an openly justified reason whose entitlement derives from a complex series of inferences generated from John's present belief-value set. Finally, Case C illustrates an openly justified reason whose entitlement derives from its ability to be soundly inferred from John's belief-value given additional, though easily accessible, information. I take Cases A and B to be relatively simple cases of bona fide reasons for action. Case C, however, raises concerns since the reason is not presently recognized by John. However, the reason is one that he can rationally recognize given an additional amount of information. In this way, John can be said to have a reason not to push the red button. I submit that this is an intuitive use of the idea of "having a reason". It makes perfect sense to say that John has a reason not to push the red button even if he is presently unaware of the button's function.

The notion of recognition is often ambiguous in public reason liberalism and contractualism. Actual or "reasonable" recognition is run together with rational recognition. Open justification can give substance to liberal practical reason's idea of recognizing a reason. Specifically, a reason can be recognized by John when it is openly justified to John. Let me state here clearly that it is the notion of open justification that grounds the notion of rational recognition to which I appeal throughout the dissertation. From here on out, whenever I claim that John has a reason, that he recognizes a

reason, that he is rationally committed to a reason, that he can reasonably be expected to recognize a reason, that he accepts a reason, that a reason is acceptable to John, and so on, I mean that the reason is openly justified to John or at least that members of the public take the reason to be justified to John. In other words, John is entitled (or publicly thought to be entitled) to affirm the reason. In taking himself to have such a reason, John does no wrong. And he does no wrong because the reason is either generated directly from percepts and feelings or by means of sound inference from his present or added information.

II.ii Justificatory Reasons Defined

I take a reason to be justificatory whenever it is the sort of reason that can potentially play a direct role in public justification, that is, it is at least a ground for the justification or defeat of the grounds of interference.⁴⁵ Given that public justification relies on open justification, I define a justificatory reason as follows:

Justificatory Reason: X is a justificatory reason for A if and only if X is openly justified to A.

Note that open justification entails that all justificatory reasons are indexed to agents. A justificatory reason is always a reason *for* some individual. It is a reason that is internal to an agent's belief-value set, internal to her own mental life. For purposes of illustration, we might contrast this conception of reasons with that of Steven Wall's externalist conception of reasons.⁴⁶ We can understand a reason as external for John when it is a reason for John regardless of whether it can be openly justified to John. Wall's view follows from his "Universality Principle":

If there is a moral reason for agent A to do x in circumstances C, and if agent B is in circumstances no different in all relevant respects from C, then there is a moral reason for B to do A.⁴⁷

For Wall, if John has a reason to X then *ipso facto* everyone has a reason to X in the same circumstances. Reasons are indexed to circumstances, not to those who find themselves in such

⁴⁵ We may have external reasons that are not openly justified by definition.

⁴⁶ Wall 2010, pp. 123-151 and Wall 1998, pp. 63-104.

⁴⁷ Wall 2010, p. 134. While Wall speaks of "moral reasons" I apply the same ideas to practical reasons generally.

circumstances. An external reason applies to John only if the reason would apply to anyone in John's circumstances. To illustrate, suppose that there is an external reason that requires husbands to be kind to their wives. If John and Reba are married, then John has a reason to be kind to Reba merely because there exists an external reason that applies to husbands and since John is a husband, the reason applies to him. Wall's externalist view breaks the link between reasons to act and rational commitment. Recognizing my reasons is not a matter of reasoning from my belief-value set. Wall's conception of reasons works from facts about the world to what I have reason to do. A theory of justificatory reasons, in contrast, begins with an individual's mental primitives.⁴⁸ I am not claiming that *all reasons* must so begin. There may indeed be external reasons. But all reasons relevant to liberal practical reasons must be internalist in terms of open justification.

II.iii Public Reasons Defined

I have thus far spoken of *justificatory* reasons rather than *public* reasons. But the idea of a justificatory reason helps to describe the nature of public reasons. We cannot take all justificatory reasons to be public, as some openly justified reasons seem somewhat private. They might be reasons that others do not recognize as reasons in any sense of the term. I take a justificatory reason to a public reason when the justificatory reason is *intelligible* to others. I will intelligibility in more detail below (Ch.5, I.ii). But for now, I claim that a justificatory reason for A is intelligible if and only if members of the public P regard the reason as openly justified for A according to A's evaluative standards. Members of the public must be able to see John's reasons *as justificatory for John*. Otherwise, it is unclear how the reasons can enter into the process of public justification. If John's reasons to reject a proposal do not look like reasons *at all* then how can members of the public be expected to accept the reason as a defeater? The converse raises the same worry: If John assents to a norm based on considerations members of the public regard as unintelligible, how can they know they are treating him as free and equal? John's evaluative standards are (mostly) John's standard of belief entitlement, the standard he uses to evaluate his commitments, values and beliefs. I will assume in Chapter 5 that reasonable pluralism holds for evaluative standards among members of the public. In other words, they can disagree about what evaluative standards are valid. Much of the time, the set of justificatory reasons will overlap with the set of public reasons. But this will not always be true. In some cases,

⁴⁸ This distinction in some ways maps onto Stephen Darwall's distinction between doing moral theory from the "inside out" and the "outside in". See Darwall 1986.

members of the public may not have sufficient information to regard John's reason R as meeting his own evaluative standards even if *in fact* R meets his evaluative standards. In those cases, John will have some justificatory reasons that cannot be employed in the process of public justification. We might imagine, for instance, that John takes himself to have an openly justified reason for action based on a set of evaluative standards that would be opaque to even quite rational and well-informed members of the public. In such a case, John would have a justificatory reason that could potentially count as public, but whose publicity would be blocked by its unintelligibility. Note also that John can have justificatory reasons that are intelligible to only *some* members of the public, perhaps within his particular religious community or even within his family. In that case, there may be moral demands that can be enforced only by friends because they have access to John's commitments that members of the public at the right level of idealization lack. Thus, moral justificatory practices can take place within groups smaller than the public.⁴⁹ Justificatory reasons can thereby establish justified norms for groups within the public that do not extend to the public themselves. The dissertation is only worried about public reasons, however, so let us tentatively define a public reason as follows:

Public Reason: X is a public reason for A if and only if it is (i) intelligible and (ii) a justificatory reason.

My definition of a public reason holds that having an intelligible justificatory reason is both necessary and sufficient for having a public reason. Remember that a public reason is a reason that can justify interference and coercion. Coercing or interfering with others on the basis of public reasons is required to treat others as free and equal, to treat them with dignity and respect.

My definition of a public reason raises an important complication that must be addressed right-off. There appears to be a tension between the two individually necessary and jointly sufficient conditions for being a public reason. The tension can be brought out in two cases: (a) where a public reason is intelligible but not justificatory and (b) where a public reason is justificatory but not intelligible. Consider two cases. In Case D, John has an openly justified reason X but members of the public do not see X as openly justified to John given John's evaluative standards. Nonetheless, members of the public coerce John on the basis of X. For the public reason liberal, coercion is only compatible with the freedom and equality of persons if it is based on a public reason. So we must ask whether members of the public treat John as free and equal when they coerce him on the basis

⁴⁹ I thank Keith Hankins for this point.

of X. The answer seems to be no. In case D, members of the public coerce John based on considerations they do not regard as reasons for John even by John's evaluative standards. They know that they are coercing John for reasons they believe he does not recognize. This behavior seems incompatible with the spirit of the Liberal Idea as members of the public fail to reason from John's standpoint. Thus, intelligibility seems to be necessary for a reason's being public.

But consider Case E. In Case E reason X is *not* openly justified to John, but members of the public believe that it is. They regard reason X as openly justified to John. Now, suppose that members of the public coerce John on the basis of reason X. Do they treat him as free and equal? There seems to be a case either way. On the one hand, members of the public demonstrate a willingness to coerce John only on the basis of what they take to be his reasons.⁵⁰ On the other hand, they have in fact failed to coerce John on the basis of openly justified reasons. In other words, John has a complaint that he is not being treated as free and equal because he is being held to a norm that he does not endorse. The aim of public reason liberalism is to produce a society where all are free, not merely where everyone believes that everyone else is free. And what it means to be free on public reason liberalism, as I understand it, is that everyone's freedom is preserved because they are coerced only on the basis of their reasons.

Honestly, I am not sure what to say in this case. But it seems to me that Case E forces us to recognize that it is possible to treat others as free and equal by coercing them on the basis of reasons that are not openly justified to them. Thus, open justification seems to be an unnecessary standard. While it is true that intelligibility includes reference to open justification (members of the public must believe that they are treating John according to openly justified reasons), the reasons they take to be openly justified need not *be* openly justified. In this case, I can only give my view and move forward. Open justification serves as a regulative ideal within public reason liberalism. Members of the public treat one another as free and equal when they coerce one another only on the basis of intelligible reasons. And so it seems best to understand reasons as public when and only when they are intelligible. However, if the Rousseauian ideal of public reason holds, then members of the public must aim to treat others according to reasons that are *in fact* openly justified. The telos of public reason liberalism is to establish a set of norms and laws that are in fact openly justified to all members of the public, even though they can treat one another as free and equal (and thus do no

⁵⁰ Let us assume in Case E that members of the public would *not* coerce John based on X if they did not take X to be a reason for him given his evaluative standards.

intentional wrong), when they treat one another in accord with merely intelligible reasons. So in light of this, let me define two types of public reasons:

Exemplary Public Reason: X is an exemplary public reason for A if and only if it is (i) intelligible and (ii) a justificatory reason.

Adequate Public Reason: X is an adequate public reason for A if and only if it is intelligible.

An exemplary public reason is a reason that meets the ideal of public reason, that is, it is a reason that members of the public recognize as a justificatory reason and that is in fact a justificatory reason. A publicly justified polity is exemplary when all coercion is justified according to exemplary public reasons. To put it another way, an exemplary polity is one where intelligible reasons and justificatory reasons always and everywhere match. But the dissertation requires only adequate public reasons. So for the rest of the dissertation I shall refer to adequate public reasons as “public reasons” and leave aside the complex case of exemplary public reasons and their relations to adequate public reasons.

Before moving to the next sub-section we must consider another objection. Given the definition of public reasons on offer, John and Reba will inevitably have distinct public reasons. Some will find this suggestion absurd on its face. How can a reason be *public* if it does not apply to all individuals? I take a reason to be public for John when members of the public can coerce John on the basis of the reason and still treat him as a free and equal member of the public. In this way, a reason is public because it makes coercion and interference by members of the public permissible. The need for public reason liberals to hold that public reasons apply to all citizens push them to conceive of public reasons as essentially shared. Among the dissertation’s central points is that this is simply the wrong way to think about the relationship between reason and morality. It is because our reasons shape the conditions under which a social morality may be justified that they are public. In one way it may be better to speak of a *public use* of reasons with regard to the justification or defeat of a social-moral rule. In this way, *norms, rules and proposals* are the property of the public, not reasons.

II.iv Justifying Interference

Given that public reasons may differ, how can they justify interference and coercion? To answer this question, we must first recognize that John's public reason R cannot by itself justify interfering with Reba. If John has a public reason to interfere with Reba, he does not thereby give her a reason to comply with his interference. Instead, to justify his interference, John must appeal to a public reason *of Reba's* in order to show that she should comply. To put it another way, moral demands can properly be imposed on others only in light of two public reasons: (i) the demander's public reason to make the demand and (ii) the subject of the demand's public reason to comply with the demand or action in question. The same holds for justifying coercive actions.

Consider Wall's view for contrast. For Wall, John's external reasons can justify interfering or coercing Reba even if she lacks a rational commitment to such coercion. So long as John has "full justification" for his beliefs about the external reasons that apply to him, he can justifiably act. Wall's conception of full justification is the combination of subjective and objective justification: "... an action is fully justified, if and only if, it is both subjectively and objectively justified."⁵¹ For Wall, an action is subjectively justified "if it is an action that is required or permissible given what he or she knows."⁵² Wall's conception of subjective justification seems to take the form of internalist externalism. An action is objectively justified for Wall "if it is in accordance with right reason."⁵³ Objective justification is straightforwardly externalist. One can only be justified in acting if she acts for the right (external) reason. To illustrate, Wall describes a case of disagreement about the permissibility of euthanasia.⁵⁴ Suppose that Jill has a subjectively justified belief that euthanasia is permissible but that Jack, even if he were fully rational, would not accept the permissibility of euthanasia. For Wall, Jill is morally permitted to force her policy on Jack despite the fact that he has no access to a reason to comply: "if Jill believes that the immorality of euthanasia justifies coercive intervention ... then the fact that Jack cannot rationally accept this either also should not matter."⁵⁵ But liberal practical reason maintains that for Jill to permissibly interfere with Jack, she must not only be entitled to believe that her demand is justified to Jack, but Jack must be rationally committed

⁵¹ Wall 2010, p. 134.

⁵² Wall 1998, p. 102.

⁵³ Ibid.

⁵⁴ Wall 2010, p. 124.

⁵⁵ Ibid., p. 139.

to complying with the demand. In contrast, Wall suggests that a citizen should not reason from the standpoint of others, but instead “reason from his own standpoint.”⁵⁶

II.v Summary

In sum, a public reason, the unit of moral justification within liberal practical reason, should be understood thusly:

- (i) R is a(n) (adequate) public reason for A iff it is intelligible.
- (ii) An interferer A’s public reason to interfere with B is not sufficient to justify interference with B; B must have a public reason to comply with or permit the interference.⁵⁷

III: Second-Person Reasons

Public reason liberalism requires that we reason from our own perspective *and* from the perspective of others. As such, any conception of liberal practical reason must appeal to two sorts of reasons—reasons that derive from our own perspective and reasons which derive from taking the perspective of others. Darwall is well-known among moral philosophers for arguing that there is a special kind of moral reason, a *second-person* reason that derives from taking the standpoint of others. It is the task of Section III to make sense of the idea of a second-person reason in light of the criticisms lodged against Darwall’s conception (Ch.3, III.iv). The notion of a second-person reason will help to make sense of the normative force of claims made by public reason liberals against the integrity-based complaints of their critics.

III.i: Second-Person Reasons, Second-Person Rules

With a basic picture of public reasons and an account of moral rules, we are now in a position to reconceive of the notion of a second-person reason. Liberal practical reason understands all public reasons as originating with one’s rational commitments. On this, Darwall concurs: “second-personal

⁵⁶ Ibid., p. 137.

⁵⁷ Note that included among those who interfere are those who make moral demands and those who are interfered with include those on whom such demands are made.

address is always also first-personal.”⁵⁸ That said, Darwall denies that second-person reasons are “merely first-personal.”⁵⁹ The Darwallian conception of second-person reasons requires these reasons to somehow bridge the gap between different first-personal perspectives. They are therefore first-person “plural” reasons. In light of the arguments of Chapter 3, I shall offer a different view: second-personal reasons are public reasons that concern moral rules and whether those rules are publicly justified. I must stress that this modification of Darwall’s view is *absolutely vital* to the fundamental problem of the dissertation. In my view a lack of precision concerning the structure of moral justification has led to needless conflict between public reason liberals and integralists. Unless we understand precisely what a second-person reason is, we cannot understand the conflict between public reason liberals and integralists. And if we cannot understand the conflict, we certainly cannot find a solution to it.

Chapter 3 demonstrated why moral rules are fundamental for generating a justified moral order. Individuals cannot satisfy the informational burdens required to know why moral rules are justified to all. Like prices in the market, moral rules are required to aggregate justificatory reasons and solve moral coordination problems. In doing so, moral rules obviate the need for dyadic or “plural” perspective-taking in order to generate moral authority. If we understand all public justification as occurring by means of moral rules, we can clarify the process of public justification and the relationship between publicly justified morality and integrity. This is not to say that the kind of reciprocal recognition that Darwall discusses is not required for moral authority or the justification of interference and coercion. It is merely to say that the proper understanding of how reciprocal recognition generates moral authority must include an account of moral rules.

To see this, note that the presence of moral rules allows us to usefully distinguish among two types of public reasons: (i) those that directly concern moral rules and (ii) those that do not. I shall call the former “second-person” reasons. Second-person reasons are second-personal because they concern moral rules that are justified (or unjustified) to members of the public. But second-personal reasons can be further distinguished in how they relate to moral rules. I take justified moral rules to generate three types of second-person reasons: (i) reasons to interfere in accord with the rule, (ii) reasons to permit the interference and (iii) reasons to comply with the interference. To illustrate, consider the following justificatory situation: John Breckenridge and Reba Hollingsworth are members of the Southern aristocracy in the American South during the mid-19th century. In this

⁵⁸ Darwall 2006, p. 10.

⁵⁹ Ibid.

culture, it is considered an affront to a woman for a man to keep his hat on in her presence. Suppose that the male-hat-removal rule is publicly justified to both John and Reba. Assume further that John doesn't much care for Reba and thus when she walks in the room, he rudely refuses to remove his black stove-pipe hat. In this case, three public reasons arise: (i) Reba has a public reason to demand that John take off his hat (though perhaps she is too proper to make such a fuss), (ii) John has a public reason to permit Reba to make the demand and (iii) he has a public reason to comply with the hat-removal rule. John and Reba are in a justificatory situation where Reba has justified authority over John to require he comply with her demands. Since the rules are justified to both parties and they recognize this fact, Reba can exercise her moral authority over John by means of the rule. However, in cases of unjustified rules, this edifice of reasons disappears. Suppose that we replace the traditional Southern John Beauregard with John Stuart Mill, a liberal feminist despite his membership in the British Aristocracy. John Mill believes that men and women should be treated equally and that hat removal rules reinforce a norm of female submission that degrades their equal status. Therefore the male-hat-removal rule is not publicly justified to John Mill.⁶⁰ It follows that Reba lacks a public reason to demand that John follow the rule and John has no public reason to permit Reba to make such a demand of him and less reason still to comply with it.

We may further refine these types of public reasons by considering how they relate to the thin account of publicity developed in Chapter 3. Members of a moral community need only know *that* some moral rule is publicly justified to all, but not why.⁶¹ Therefore, if Reba has a reason to demand that John Breckenridge take off his hat, her reason may be as simple as "Gentlemen remove their hats in the presence of a lady." That the rule is justified and binding is sufficient reason to interfere. If Reba wishes to make the demand of John and he objects, she need merely cite the rule. No additional reason is required. However, a person can comply or reject compliance for many reasons. The reasons John Breckenridge has to remove his hat in Reba's presence will shift in accord with his reasons and values. These diverse reasons must constitute objections to the rule or to Reba's actions with respect to it. All second-person reasons interface with the rule. Public reasons do not "reach out" and create obligations, as Darwall's second-person reasons attempt to do. Instead, moral rules form normative bridges between individual perspectives. All legitimate moral interaction must occur by means of the bridge. Public reasons acquire their second-personality in virtue of being hooked to a moral rule.

⁶⁰ I assume, in light of II.iv, that at the right level of idealization, Reba recognizes that John has no intelligible reason to comply.

⁶¹ I will discuss how we know when rules are justified in Ch.10.

III.ii Second-Person Reasons Defined

We may now define second-person reasons:

Second-Person Reason: X is a second-person reason for A iff (i) X is a public reason for A and (ii) X is a reason to interfere, permit interference or comply with interference in accord with some moral rule M intelligibly believed by A to be publicly justified.⁶²

The class of second-person reasons is a subset of the class of public reasons, thereby substantiating condition (i). Second-person reasons are that subset of public reasons that concern the initiation of interference or a response to interference in accord with rules. Second-person reasons are only bona fide if they are based upon some moral rule M, that is, they are reasons with respect to M to issue a demand to comply with M, to object to M, etc. The person who has the second-person reason must also believe that M is publicly justified. If she did not so believe, she would have no reason to issue demands based upon M or to permit or comply with interference justified by M.⁶³

III.iii Second-Person Reasons Refined

We can now refine the notion of a second-person reason and its relationship to a rule with a concrete example. In *The Picture of Dorian Gray*, Oscar Wilde tells the story of a beautiful young man, Dorian Gray, who adopts a hedonist philosophy of life from Lord Henry “Harry” Wotton, due to Lord Henry’s wit and influence. The two men had been introduced by an artist, Basil Hallward, who was so infatuated with Dorian’s beauty that he painted a portrait of him as a result. One day Lord Henry tells Dorian that all beautiful things grow corrupt with time and that even Dorian’s beauty will fade. Horrified, Dorian wishes that his portrait will age instead. His wish comes true, but it also corrodes in response to Dorian’s sins. And with every sin and every year, Dorian’s portrait becomes increasingly hideous, while Dorian’s appearance remains unchanged. Over decades of indulgence, Dorian keeps the portrait hidden. But one day, Basil discovers the painting and encourages Dorian to repent for his sins. In response, Dorian blames Basil for his fate and murders him.

⁶² Recall that rule M is publicly justified to A when A has a conclusive reason to endorse M, that is, a public reason that defeats all competitor reasons in her system.

⁶³ Of course, we cannot rule out cases where Reba has some independent reason to make demands that just so happen to be in accord with rule M.

When Basil discovers Dorian's portrait, he exclaims:

Good God, Dorian, what a lesson! What an awful lesson! Pray, Dorian, pray. What is it that one was taught to say in one's boyhood? "Lead us not into temptation. Forgive us our sins. Wash away our iniquities." Let us say that together. The prayer of your pride has been answered. The prayer of your repentance will be answered also.⁶⁴

Let us say that Basil interferes with Dorian when he demands that Dorian repent. The moral rule which Basil demands that Dorian follow is "Repent for your sins" which I will presume was publicly justified in the late Victorian culture in which *The Picture of Dorian Gray* took place (though, ironically, not to Wilde). I will now diagram the second-person reason Basil had to make the demand in line with the repentance rule. I shall represent the elements of the reason as follows: "B" = Basil, (b) "D" = Dorian, (c) "M" = the repentance rule, (d) "A" = action, and (e) "T" = time. We may decompose Basil's second-person reason to demand as follows:

Second-Person Reason to Demand

R is a second-personal reason for B to demand D follow rule M iff:

- (i) R is intelligible to B;
- (ii) R has M as a content;
- (iii) B intelligibly believes:
 - a. Rule M is publicly justified to B's community;
 - b. Action A violates M;
 - c. Community member(s) D is A-ing at time T;
 - d. T_x is appropriately close to the present T;

Condition (i) refers back to the claim that all second-person reasons are public reasons, since a public reason for B must be intelligible to B. Condition (ii) requires that reason R be a reason because of rule M. A second-person reason for Basil to Demand that Dorian repent might be: "In Dorian's circumstances, repentance is required." This reason R refers to a rule M requiring

⁶⁴ Wilde 1992, p. 125.

repentance when reason R's normative force is recognized. If Basil cites this reason when demanding that Dorian repent, he thereby refers to rule M.

Condition (iii) requires several implicit or explicit beliefs. Note first that all four of these beliefs must be intelligibly justified. If they are not intelligible the reason generated by the conditions cannot be *bona fide*. Why must these beliefs actually be intelligible rather than *in fact* openly justified? If Basil acts on reasons based on beliefs that members of the public think groundless, they will regard him as acting disrespectfully, as failing to treat Dorian as free and equal when he demands that Dorian comply with the rule of repentance. Basil will still fail to treat Dorian as free and equal if members of the public believe that Basil has no relevant openly justified beliefs. Thus the beliefs specified by condition (iii) need public recognition like the reasons that bind the community to the rule. In general, it seems that I should not act on beliefs I am not entitled to hold but that I should still carry myself insofar as I can in accord with reasons based on intelligible beliefs. R is a second-person reason when it is intelligible as are the beliefs upon which it is based in condition (iii).

Condition (iii-a) requires that Basil believe that rule M is publicly justified to Basil's community.⁶⁵ Basil must believe that rule M is conclusively justified, for if he did not, his demand would be unjustified since he recognizes that others may have defeaters for M. Also note that (iii-a) refers to Basil's community. The relevant rules may be justified to different communities. Basil might demand that a family member follow a rule justified to members of his household, or he might require that a fellow citizen obey national traffic laws. The rule might even be universal: Basil may require that a foreign diplomat not steal from him. Reference to community is necessary because it specifies who has the standing to make demands of others in line with the rule.⁶⁶ If rule M is only justified to Basil's community, then only members of Basil's community can demand that other community members follow rule M. Thus, condition (iii-a) indexes appropriate standing.

Conditions (iii-b), (iii-c) and (iii-d) are simpler. (iii-b) requires the openly justified belief that a given action violates a rule. (iii-c) is an observational belief, where Basil identifies Dorian's actions as belonging to a class of actions A. (iii-d) is a time indexing belief, which contains a subtlety. Suppose that John is haranguing Reba in a public library where Sarah is present. A year passes and Sarah runs

⁶⁵ Recall that public justification requires conclusive reasons. I explored the conclusiveness requirement in Ch.1, III.iii. A conclusive public justification of rule M shows that the reasons to endorse M defeat reasons to embrace any other rule. Thus, a conclusively justified rule is one that an individual has decisive reason to follow.

⁶⁶ Within public reason liberalism, members of the public are the relevant community.

into John on the street. Were she to then demand that John not harangue his wife, her demand would fail to be appropriately close to the time when the observational belief (iii-c) was true.

If conditions (i), (ii) and (iii) are met, then reason R is a second-person reason in Basil's system. R has second-personality in two senses: (a) it relates to demands made upon others and (b) it is a reason to act in accord with a publicly justified moral rule. Finally, notice that the definition of Basil's reasons specifies numerous occasions for defeat. If Basil's beliefs (iii-a) through (iii-d) are unintelligible, his second-person reason is defeated. Defeater reasons vary in type (see Sec. V). Before continuing, I must emphasize that the beliefs mentioned here will rarely be occurrent in an individual's mind when making a moral demand. For Basil to rightly demand that Dorian repent, he needs only believe that a failure to repent in Dorian's circumstances would be blameworthy. In so believing, Basil in effect holds that rule M requiring repentance is one justified to members of his community, that Dorian is a member of the community and that he is presently violating the rule.

There are many other types of second-personal reasons: reasons to comply with demands, reasons to interfere, and reasons to comply with interference, reasons to reject interference, reasons to block interference, reasons to propose interference, reasons to permit interference, etc. But all of these reasons concern rules that must be intelligible to a community. The great advantage of understanding second-person reasons in relation to moral rules is two-fold: first, moral rules allow us to neatly separate second and first-personal reasons and second, moral rules allow us to neatly distinguish between types of second-person reasons while showing that they retain a common rule-regarding core. We will see below that the second-personality of liberal practical reasons will provide an attractive model of understanding what makes a reason public within public reason liberalism, that it is a reason relevant to the public justification of a coercive action or set of actions, not that it is shareable, accessible, and so on. In this way, a large assortment of reasons can count as public, including many religious reasons. The idea of a second-personal reason helps to construct a theory of liberal practical reason with a place for religious reasons. The theory can satisfy integralists while simultaneously providing a more attractive account of public reason liberalism.

IV: First-Person Reasons, Ideals and Integrity

Our account of liberal practical reason also includes the idea of first-person reasons which bear a complex relation to second-person reasons. First-person reasons explain the normative force of

complaints issued by integralists. However, the complete model of first and second-person reasons will deprive neither source of normative concern of their force.

IV.i First-Person Reasons

First-person reasons are not rooted in the practice of social morality as directly as second-person reasons. A necessary condition for a first-person reason is that it is not by itself a reason to interfere, permit interference or comply with interference in accord with a social-moral rule. Specifically, first-personal reasons lack reference to second-person rules in their content in contrast to second-person reasons. They are not reasons issued in accord with a social-moral rule. In this way, first-personal reasons need not have any interpersonal content. For instance, John might have a first-person reason to watch another hour of TV. His action makes no reference to a rule of interpersonal morality. It is simply a public reason to engage in some pursuit or activity.⁶⁷ We will see below that first-person reasons tend to be related to an individual's goals, projects, plans, that is, his integrity. First-person reasons are reasons agents have to achieve their goals and be loyal to their principles (or reasons involved in achieving those goals or remaining loyal to those principles).

That said, first person reasons are still relevant to public justification because they can defeat or affirm moral rules. For instance, Reba may have a first-person reason to build a sand castle. But as a result of the presumption against interference and the Public Justification Principle, Reba's reason may be able to defeat John's reasons to interfere with her. For her reason shows that the rule to which John appeals is not publicly justified. In this way first-person reasons help determine whether a moral rule is justified. Remember: second-personal reasons are only reasons if the rule they concern is publicly justified. And the rules can only be publicly justified if they are (to a larger extent) compatible with the first-person reasons of members of the public. Thus, second-person reasons are only reasons with respect to publicly justified moral rules which are in turn publicly justified in accord with first-person reasons.⁶⁸ Consequently, if a first-person reason shows a moral rule to be unjustified, then it shows that there are no second-person reasons to interfere on the basis of that rule. There is therefore a *dependence* relation between second-person and first-person reasons concerning demands made in accord with those rules. Second-person reasons are only reasons *when*

⁶⁷ Of course, John may have non-public reasons to act that are related to his ideals and goals. But these reasons cannot enter into public justification.

⁶⁸ And in accord with other second-person reasons.

first-person reasons ground them. The class of first-person reasons helps to demarcate the class of second-personal reasons by determining which moral rules are justified or not.

IV.ii Individual Ideals

To account for first-person reasons, I turn to a little-known article of P.F. Strawson's, "Social Morality and Individual Ideal" where Strawson articulates a precursor conception of the first-second personal tension we are exploring. Strawson contrasts "individual ideals" with "social morality." For Strawson, individual ideals are those principles and evaluations that "govern the choices and decisions which are of the greatest importance to men."⁶⁹ Individual ideals do not demarcate the entire set of an individual's reasons. Instead, they provide reasons of the "greatest importance." Strawson locates these ideals within the "region of the ethical" rather than the "moral" and argues that they "often take the form of general descriptive statements about man and the world."⁷⁰ These ideals are primarily descriptive, not prescriptive, though they are frequently employed to generate prescriptions.⁷¹ Strawson claims that these ideals:

[C]an be incorporated into a metaphysical system or dramatized in a religious or historical myth. Or they can exist—their most persuasive form for many—as isolated statements One cannot read Pascal or Flaubert, Nietzsche or Goethe, Shakespeare or Tolstoy, without encountering these profound truths In most of us the ethical imaginations succumb again and again to *these* pictures of man and it is precisely as truths that we wish to characterize them while they hold us captive.⁷²

Individual ideals help define the purpose of human life and how humans live best, but they are not themselves first-personal reasons. Instead, they are used to *generate* reasons. If Reba's individual ideal is Medieval Roman Catholicism, then she sees humanity's highest good as the contemplation of God. So if Sarah wishes to follow her individual ideal, she must engage in particular projects and adopt certain principles in order to do so. Her aim of following her ideal thus gives her a number of

⁶⁹ Strawson 1974, p. 31.

⁷⁰ Ibid.

⁷¹ Sometimes individual ideals are rarely used to generate prescriptions. As Strawson notes, individual ideals "may receive the scantiest expression in their actual behavior and would call for the most upsetting personal revolutions if they received more." Ibid.

⁷² Ibid., p. 32.

reasons of the “greatest importance” such as reasons to celebrate the Mass or attend confession. These first-person reasons lie at the core of our system of mental primitives, our belief-value set. They structure and generate reasons outside of the core. For instance, to celebrate Mass effectively, Reba will want to know the liturgy of the Mass and will therefore have a reason to commit it to memory. If Reba wishes to faithfully confess her sins, she must engage in a self-inventory of her sinful actions. She therefore has a reason to make an honest self-examination. These more peripheral reasons are first-personal as well, but they depend upon Reba’s core first-person reasons that are generated by her commitment to pursuing her individual ideal.

IV.iii First-Person Reasons Defined

With the notion of an individual ideal in hand, we can now define first-person reasons:

First-Person Reason: X is a first-personal reason iff (i) X is a public reason and (ii) acting on X is a means toward acting in accord with a full or partial individual ideal.

First-personal reasons are public reasons, so condition (i) is clear. But condition (ii) raises some complications. First, it is necessary to define the notion of a partial individual ideal. I take a partial individual ideal to be a worldview or personal ideal that is incomplete in that it covers only certain aspects of life or has relatively vague content. A partial individual ideal must still consist in a picture of how a life is well-lived, but it can be relatively vague, as vague, in fact, as “being happy”. First-person reasons should relate to acting in accord with an individual ideal, but by defining these reasons as such, we leave open the possibility that there are public reasons that are neither first-personal nor second-personal. For instance, John may have a reason to scratch his ear or to take a shower rather than a bath that is not related to his individual ideal in any direct sense. But we can set these reasons to the side, as they are not central to our conflict. First-person reasons must be a means toward acting in accord with an individual ideal. Accordingly, first-personal reasons may be core, such as Reba’s reason to learn how to contemplate God, or they may be peripheral, such as Reba’s reason to memorize the liturgy of the Mass. But both reasons, core and peripheral, are means towards achieving or living up to Reba’s individual ideal. In this way, even minor, peripheral reasons for action can achieve first-personality by being means towards acting on relatively less peripheral

reasons. For instance, Reba may have a first-person reason to get up at 5:00 am rather than 6:00 am so she will have time to say her prayers before breakfast is served in her monastery.

I noted above (IV.i) that first-person reasons bear dependence relations to second-person reasons. In these cases, it may often appear that a reason can be both first and second-personal at the same time. For instance, Reba's commitment to Roman Catholicism may provide her with reason to object to moral rules forbidding the discussion of religion in the public square. Reba thereby seems to have a reason that is both first-personal, because it is a means toward acting in accord with her individual ideal and second-personal, because it is a reason relating to interference in accord with a moral rule. In this way, there seem to be public reasons that have both first and second-personality. I see no reason to disallow the possibility of such an overlap but in many cases it is better to understand such reasons as two distinct reasons defined by their relation to other reasons and rules. So, in this case, Reba has a reason to object to the privatization of religious belief. I think it is clearer to say that she has a first-person reason to create a social space for the expression of her Catholic beliefs (which is part and parcel of living up to her individual ideal of Roman Catholicism) and a second-person reason to reject the privatization norm (which relates to the rejection of the legitimacy of moral demands to privatize belief in accord with the privatization rule). Whether we describe the reasons Reba has to object to the privatization norm as one reason or two reasons is of interest, but for now we must set this case aside. For the remainder of the dissertation, I will simply treat reasons that appear to have both first and second-personal aspects as two similar but distinct first and second-personal reasons.⁷³

IV.iv Social Morality and Individual Ideal⁷⁴

We can now specify how first and second-person reasons connect. Not only are second-person reasons dependent upon first-person reasons, first-person reasons can be more easily acted upon

⁷³ Much of the ambiguity depends upon how reasons are individuated. I think that reasons are individuated by intentions and that, accordingly, the ways reasons are best defined partly in terms of the relations they bear to intentional states. This will make providing a satisfying philosophical analysis of the individuation of reasons a difficult matter. It is for this reason—and the fact that little seems to hang on the question—that I set the matter of reason individuation aside.

⁷⁴ Some might worry that by allowing first-person reasons to sometimes undermine second-person reasons that second-person reasons will lose their ability to override first-person reasons. This is not so, but to explain why second-person reasons are often overriding involves, in effect, answering the “Why be moral?” question. For now, I assume that our moral psychologies are such that we frequently take reasons of right (interpreted second-personally) to override reasons of the good (interpreted first-personally). This dissertation outlines one method of showing how the two types of reason congrue or match but it assumes at the outset that sometimes reasons of right trump reasons of good.

within a social morality that is sustained by second-person reasons.⁷⁵ Individual ideals can only be realized within social morality, since “the ideal images of which [Strawson] spoke demand for their realization the existence of some form of social organization.”⁷⁶ For Strawson, this is true in both a constitutive and empirical sense. In the constitutive sense, some ideals can only be realized within a concrete form of social organization by definition: John’s ideal may be to become President of the United States, a goal that essentially involves social morality. One can only be the President of the United States in a world with a complex network of moral rules. Otherwise, institutions like the Presidency could not exist. In the empirical sense, the development of individual ideals requires social cooperation and the moral rules that make social cooperation possible. That humans have had enough time and resources to even develop the complex individual ideals of the present day is an extraordinary achievement. Social cooperation must have already existed for thousands of years before individual ideals of any complexity could develop. But social cooperation requires social morality, moral rules enforced with blame and sanction. Thus, individual ideals cannot even arise without social morality.

Strawson’s conception of social morality resembles Baier’s conception of moral rules. First, social institutions of any kind only exist when “certain expectations of behavior on the part of its members should be pretty regularly fulfilled.”⁷⁷ For a society to exist, it must enforce behavioral expectations. Strawson terms social expectations to follow social rules “the minimal interpretation of morality.” He claims that,

[The minimal interpretation of morality] represents [social morality] as what might literally be called a kind of public convenience: of the first importance as a condition of everything that matters, but only as a condition of everything that matters, not as something that matters in itself.⁷⁸

To attain “everything that matters” the rules associated with the minimal interpretation of morality must be regularly followed. These rules are akin to the conception of moral rules developed in Chapter 3, for they are “socially sanctioned demand[s] made on an individual” that are advanced based merely upon “his membership of the society in question, or in virtue of a particular position

⁷⁵ We might have unjustified social-moral practices and thus have a social morality not based on second-person reasons.

⁷⁶ Strawson 1974, p. 33.

⁷⁷ Ibid.

⁷⁸ Ibid.

which he occupies within it or a particular relation in which he stands to other members of it.”⁷⁹ Again, moral rules are rules of social behavior that establish expectations of compliance by members of the relevant community. When individuals fail to conform, they are threatened with the prospect of “socially sanctioned demands.” A socially sanctioned demand, for Strawson, is directly related to the moral ideas of permissions, approvals, powers and penalties: “a socially sanctioned demand is doubtless a demand made with the permission and approval of a society; and backed, in some form and degree, with its power.”⁸⁰

Further, social morality cannot exist without “reciprocal acknowledgment of rights and duties”; specifically, moral demands must belong “to a system of demands which includes demands made on others in his interest.”⁸¹ Membership in a moral community requires that individuals endorse their community’s “system of demands, to the extent of genuinely acknowledging as obligations at least some of the claims which others have on one....”⁸² Like Darwall and Baier, Strawson believes that social morality requires reciprocity of claim: membership in a moral community implies recognized relations of reciprocity. In a Kantian vein, within social morality “everyone appears, grandly but unplausibly, as a moral self-legislator.”⁸³ While Strawson regards the Kantian idea of self-legislation as extravagant, he thinks it reveals “something genuinely universal in morality: the necessary acceptance of reciprocity of claim.”⁸⁴ Thus, Strawsonian social morality is usefully understood in terms of moral rules and second-person reasons. Moral rules comprise the sphere of social morality and second-person reasons can be understood as associated with the Strawsonian system of demands that imply reciprocity of claim. For this reason, I take a social morality to be a system of moral rules supported by second-person reasons that enable the development and practice of individual ideals.

This understanding of social morality helps to flesh out the relationship between first and second-person reasons. We can see that first-person reasons depend on second-person reasons in an empirical sense, since individuals will lack the ability to consistently act on their first-person reasons unless moral rules are enforced through the issuing of second-personal demands. However, first-person reasons depend on second-person reasons in a normative sense as well. To permissibly act on first-person reasons, we not only need a system of second-personal norms but a system of *justified*

⁷⁹ Ibid., p. 36.

⁸⁰ Ibid., p. 38.

⁸¹ Ibid., p. 40.

⁸² Ibid.

⁸³ Ibid.

⁸⁴ Ibid.

second-personal norms. If no moral rules are justified, then it is unclear how acting on our individual ideals can avoid trampling on the claims and demands of others. Consequently, permissibly acting on first-person reasons implies a commitment to the validity of some set of second-person reasons. While first-person reasons do not necessarily reference moral rules or second-person reasons, they imply a commitment to the validity of second-person demands and therefore a commitment to the public justification of interference. First-person reasons therefore bear a dependence relation to second-person reasons in the same way that second-person reasons bear a dependence relation to first-person reasons. If we did not acknowledge the force of second-personal demands, then we would lack the ability to act on our first-person reasons. On the other hand, unless moral rules are justified in accord with some of our first-person reasons, there are no second-person reasons. First-person reasons and second-person reasons are inextricably tied and mutually interdependent.

V: Collision and the Possibility of Rational Freedom

We have seen that social morality and individual ideals are necessary compliments, but we must also look to other ways in which the two perspectives interact. Perhaps the most challenging relationship between the two arises when the first and second-person standpoints collide. After examining this relationship, we will draw out the consequences for public reason liberalism and the integrity objection as a whole.

V.i Collision

The tension between the two types of practical reasons comes into view when we recognize that social morality does not support individual ideals without asking anything in return. As Strawson notes, social morality necessitates “some degree of general readiness to recognize claims made upon one even when this recognition cannot plausibly be said to be in one’s own interest.”⁸⁵ Baier agrees: “[Acting on principle] involves conforming to rules even when doing so is unpleasant, painful, costly or ruinous to oneself.”⁸⁶ Social morality often requires individuals to refrain from acting on their individual ideals in order to maintain the regular observance of moral rules. In fact, this sacrifice of

⁸⁵ Ibid., p. 42.

⁸⁶ Baier 1958, p. 96.

interest is what the exchange of second-person reasons is supposed to justify. Dorian cannot act on his individual ideal of hedonism if Basil continually harps on him to repent his vices, so Dorian believes that his individual ideal gives him a reason to kill Basil. Yet society cannot function if people kill one another at will. Consequently, there exist publicly justified rules against killing that defeat Dorian's first-person reason to kill Basil. On the flipside, individual ideals often show that proposed moral rules are not justified. For instance, if a slave has an individual ideal of living a free life, his individual ideal defeats rules that require that he remain enslaved. Individual ideals push back the reach of social morality just as social morality pushes back the reach of individual ideals. In some cases, second-person reasons trump first-person reasons and in others, first-person reasons defeat second-person reasons. In this way, social morality and individual ideals demarcate jurisdictions for themselves.

V.ii Defeat

Many moral philosophers like Strawson and Baier recognize that morality and our ethical ideals conflict; embedding first and second-person reasons into a contractualist framework helps specify how the conflict works. Specifically, the collisions between social morality and individual ideal often result in *defeat*, revealing some action or demand to be unjustified. Defeaters are those conditions that “cancel, reduce or even prevent” epistemic entitlement.⁸⁷ The concept of defeaters complements the idea of “defeasibility” or the property of beliefs that permits their positive epistemic status to be reduced or revoked.⁸⁸ John Pollock and Joseph Cruz delineate two primary types of defeater: rebutting and undercutting.⁸⁹ Defeaters can apply to beliefs and reasons, but I will focus on the defeat of reasons here.

A *rebutting* defeater for a claim p (that one should X) is a reason for holding that one should not X or that one has a reason to Y that is incompatible with X.⁹⁰ If a person's first and second-person reasons collide or conflict, and the conflict yields a rebutting defeater, then this means that one of the reasons has been conclusively justified. The other reason, however, remains a reason. A rebutting defeater only shows that the former reason should be acted upon in case of conflict. To illustrate, suppose that John has promised Reba that he will take out the trash, but he would rather

⁸⁷ Suddoth 2008, Sec.6.

⁸⁸ Ibid.

⁸⁹ Pollock and Cruz 1986, p. 38.

⁹⁰ Ibid. This definition is modified.

finish editing his manuscript. John has a second-person reason to keep his promise (to comply with a recognized promise-keeping rule) and a first-person reason to finish editing his manuscript (one that we assume is intelligible). Since taking the trash out won't take long and John can edit his manuscript afterward, we can assume that his second-person reason rebuts his first-person reason. The first-person reason is still a reason. After John takes the trash out, he will still have a reason to edit his manuscript. The reason is merely displaced by John's reason to keep his promises.

An *undercutting* defeater for some claim p (that one should X) is a reason to hold that one no longer has a reason to X , though this is not to defeat the claim that one should X .⁹¹ An undercutting defeater might be generated when a second-person demand is shown to be invalid because it relies on an unjustified belief. For instance, Reba might think that John has forgotten their anniversary. When she blames John for not remembering, he reminds her that it is Monday and their anniversary is Tuesday. In this case, the fact that John and Reba's anniversary is on Tuesday undercuts Reba's reason to blame John. Undercutting defeaters can be generated in other ways as well. For example, suppose that John is a member of his society's lowest caste, yet he recognizes a reason to be treated as an equal. Consequently, he has no reason to defer to the demands of higher caste members. John's reason to demand equal respect undercuts his reason to follow the caste rules: it shows that he has no reason to follow the rule.

V.iii The Possibility of Rational Freedom

Our model of liberal practical reason is complete. Let us take stock. In Section I, we found that it is best to understand public justification as meeting a presumption against interference. The presumption against interference holds that due to our conception of ourselves as self-directing agents a presumption in favor of self-directed action must be met to publicly justify a norm. In Section II, I defined the idea of an adequate public reason as any intelligible reason. Section III defined a second-person reason as any public reason directly relevant to the justification of interference with respect to a moral rule. Section IV defined first-person reasons as reasons of integrity and individual ideal. We saw that first and second-person reasons both enable the normative force of the other. But to do so, they must demarcate each other's reach. In subsections V.i and V.ii, we have seen how first and second-person reasons collide and defeat one another, or how they draw the boundary between them.

⁹¹ Ibid. This definition is modified as well.

Social morality and individual ideals bear two relations to one another: an *enabling* relation and a *demarcation* relation. Social morality enables individual ideals to form and flourish. Publicly justified moral rules make living in accord with one's individual ideal possible. However, individual ideals generate the first-person reasons that ground second-person normativity. Without full or partial individual ideals, second-personal rules cannot be publicly justified. In this way, social morality and individual ideal have a cooperative, positive-sum relationship. On the other hand, social morality and individual ideals can only perform their characteristic functions if they demarcate each other's reach. Second-person rules often require individuals to sacrifice their interests. When individuals try to act on their ideals, second-person rules sometimes get in the way. Yet individual ideals fight back, for first-person reasons often rob second-person rules of their justificatory force by defeating them, by showing that they fail the test of public justification.

Notice that social morality and individual ideals only have enabling relations *because* they have demarcation relations. Unless second-person reasons restrain first-person reasons, there will be no publicly recognized method of resolving disputes between inevitably distinct individual ideals. As such, persons who advance these ideals will find themselves in a state of irresolvable conflict, a conflict that will make them both worse off. *Second-person reasons promote individual ideals only when they restrain them.* In the same way, unless first-person reasons support second-person reasons, second-person rules cannot be justified. Thus second-personal rules have no normative power of adjudication unless individual ideals support them. But to support rules, individual ideals must also have the power to extinguish them. Otherwise second-person rules have no normative power of their own. So in the same way, *first-person reasons promote social morality only when they restrain it.*

Social morality and individual ideals seem to be in permanent and inevitable tension. And yet together they comprise a vast network of public reasons. The public reason liberal's Rousseauian hope is that somehow all of these reasons can be unified into a coherent whole. She wants the individual to be as free within the bounds of morality as he would be outside of it. The tension amongst public reasons seems to threaten this ideal. Remember that the public reason liberal aims to justify a political order to the entirety of each citizen's rational will. If social morality and individual ideals inevitably conflict, then the rational will seems to be torn between its commitment to morality and its commitment to living with integrity and with "ideal aspiration, with the vision that captures

the ethical imagination.”⁹² Our reasons to engage in public justification conflict with the ideals that inspire us, with the beliefs, values and principles that give our lives purpose and meaning.

It is this concern that powers the integrity objection. The integrity objection is, in essence, a threat to the possibility of rational freedom. Think of it this way: The public reason liberal offers a justified social morality to members of the public. Her aim in doing so is to preserve the ability of citizens to live by their own lights. She wants to use social morality to extend and develop individual ideals. But if the correct understanding of social morality requires citizens to violate their integrity, then social morality and individual ideal are enemies rather than allies. This relation of enmity is destructive to both parties. If integrity is sufficiently violated, then first-person reasons are unjustifiably constrained. When first-person reasons are so constrained, then the second-person edifice of liberal practical reason collapses and with it the possibility of a society with a social morality justified to the rational will of all. Social morality cannot leave all as free as before. Instead, citizens remain everywhere in chains.

And yet this is implausible. Is there really no system of liberal social morality under which citizens of faith retain their rational freedom? Religious citizens embrace traditions that are already committed to some system of moral rules. If they recognize the fact of reasonable pluralism, then religious citizens themselves will recognize the need for a social morality that makes a moral peace in a pluralistic society possible. If so, then rational freedom is no lofty goal. At the least we are not entitled to conclude that a social world of rational freedom is unattainable. To be justified in despairing over the possibility of rational freedom, we must conclude that no form of liberal social morality can achieve it. Certainly we are far from being able to justify such a grave thesis. What is to be done, then? A concerted and systematic effort must be made to locate a Holy Grail liberal social morality, one that can preserve the rational freedom of all citizens, including citizens of faith.

The strategy I shall follow to locate this form of liberal social morality is as follows. We saw in Chapter 1 that the Public Justification Principle admits of many variations. Chapter 2 concluded that the standard interpretation of the Public Justification Principle indeed imposes significant integrity costs upon citizens of faith. But if there is an alternative interpretation of public reason liberalism that can avoid imposing significant integrity costs, reconciliation between social morality and individual ideal will be in reach. Such an interpretation of social morality can reasonably ask the integralist to pay the integrity costs necessary for a liberal social morality. For remember: reasonable citizens of faith cannot demand that they never refrain from acting on their first-personal reasons.

⁹² Strawson 1974, p. 46.

Clearly they have some interest in a social morality. If so, then citizens of faith can retain their rational freedom under a liberal social order. The challenge to the standard version of public reason liberalism is that it generates a tension between social morality and individual ideal severe enough to suggest that the citizen of faith's first-person reasons cannot be rebutted. The concern is that citizens of faith cannot be rationally free when they are governed by the standard conception of public reason liberalism. An alternative version of public reason that imposes fewer integrity costs will inevitably preserve the tension. Nonetheless, its advocates will be within their rights to ask citizens of faith to sacrifice. In this way, by being governed by the alternative form of public reason liberalism, the integrity of citizens of faith *will not be violated* because their first-personal reasons will be rebutted by second-personal reasons. Under this alternative version of public reason liberalism, citizens of faith maintain their freedom despite being required to refrain from acting on the full range of reasons proscribed by their individual ideals. The citizen of faith will be both restrained and free. She will self-legislate her own restrictions, just as Rousseau and Kant hoped. The integralist can raise no integrity objection to this form of social morality since the citizen of faith will impose these restraints *on herself*.

VI: Conclusion

Public reason liberals and integralist liberals represent the two opposing camps of the first and second-person. Both sorts of liberal mistrust one another and often believe that the other side acts out of bad faith. But if our theory of liberal practical reason accurately models the conflict, it shows that the role of religion problem would arise even among reasonable and rational representatives of both sides. It is my contention that the role of religion debate is simply the manifestation of a tension at the heart of liberal practical reason. The freedom to live in accord with our ethical aspirations presupposes its own limitations by social morality. There is no single way to draw a boundary between the reach of these aspirations and the social morality that restricts them. The conflict will always be with us. The point of a liberal theory of justice is to show how rational freedom is possible despite the tension. The irony of the debate is that public reason liberals, in their attempts to resolve the tension, have substantially alienated people of faith. And yet the pushback by integralists has not been well-received because they reject public reason liberalism at its root. Consequently, public reason liberals see integralists as critics of liberalism *per se*. In reality, integralists

are the representatives of the first-person point of view, the voices of those whose integrity liberalism must preserve. Both sides have a point.

To establish peace between public reason liberals and integralists, public reason liberals must be prepared to modify the standard conception of public reason liberalism. But how might this modification go? The strategy of Part II is to use the foundational second and first-personal values of public reason liberalism to select among the variations of the Public Justification Principle. Specifically, we must find a version of public reason liberalism that both extends the reach of integrity and offers a more plausible interpretation of the ideal of public reason on its own terms. The hope is that once we locate an alternative, the integrity costs imposed by the alternative will be sufficiently low to give us ground to conclude that in submitting to laws generated by the alternative, the citizen of faith retains her integrity within the bounds of public reason.

Part II

Reconciliation

Introduction: Two Metrics and Two Variables

Part I ended with a problem: first and second person reasons inevitably conflict, and yet they are both required to produce a complete picture of social morality. The conflict between public reason and integrity is similar: we reason publicly in part to make social space to live by our own lights. And yet, to reason publicly, individuals must limit their claims to act on their ideals to the boundaries set by public reason. The tension between liberalism and faith is an acute version of this conflict. Just as the first-personal/second-personal conflict threatens rational freedom, so does the conflict between liberalism and faith. The integrity objection is successful when second-person morality becomes so restrictive that the unity of liberal practical reason is shattered. While all reasonable people must be willing to compromise with others by subordinating their individual ideals to social morality, many sincere persons of faith will find certain forms of liberal social morality excessive and thereby restrictive of their integrity to an objectionable degree. Public reason liberals must care about integrity in order to make sense of public justification, and so they must be worried about the claims of citizens of faith forcefully raised by their integralist critics. So, what's a good liberal to do? Towards the end of Chapter 4, I argued that we must locate an alternative form of liberal social morality. Specifically, we must locate an interpretation of the Public Justification Principle that significantly reduces the first-personal burdens imposed by public reason. The conflict between public reason liberals and their integralist critics can be resolved when the public reason liberal can show that the integrity costs of social morality are small enough to render the rational freedom of all a realistic ideal.

The Public Justification Principle admits of many variants, so how are we to select among them? In my view, we must utilize the foundational values of public reason liberalism to select among the variants. What we need are *metrics* for deciding which interpretations of the Public Justification Principle are superior to others. A metric is a method of employing a value to select among instantiations of the value. For example, a ruler counts as a method for measuring height. It can be used to determine which individual is taller than another and thus enables the ordering of individuals by height. If we are looking for the tallest person out of a group of ten, the ruler is the metric by which we do so. To give a normative example, suppose that John is searching for the best method of bringing about more justice in the world. A theory of justice is a metric that supplies a method of figuring out the best way to implement justice in the world.

I will select among interpretations of the Public Justification Principle with two metrics: (a) a *first-personal metric* and (b) a *second-personal metric*.¹ The first-personal metric is integrity. Integrity serves as a metric when we can use integrity to determine which conceptions of public reason impose relatively greater or relatively fewer burdens on integrity. Thus integrity can be used as a metric when we can determine which conceptions of public reason impose the fewest costs on the integrity of citizens of faith. Consequently, I will select among conceptions of public reason liberalism in part by determining which conceptions impose the fewest integrity costs. This may seem to arbitrarily privilege integrity, but as we have seen, public reason liberals are committed to reducing integrity costs within the constraints of social morality. Versions of public reason liberalism that impose few integrity costs are *pro tanto* superior to versions that impose relatively more, since one *point* of public reason liberalism is to extend the integrity and liberty of citizens as far as social morality allows. The second-personal metric is reasonable pluralism. Reasonable pluralism serves as a metric when we use it to determine which conceptions of public reason express the most respect for the diverse reasoning of citizens. Remember that public reason liberalism is committed to securing the rational freedom of all. As such, versions of public reason liberalism that require that public justification obtain in light of the totality of citizens' reasons will be superior to versions that are more restrictive or that bracket the set of reasons to which norms must be sensitive. Reasonable pluralism also serves as a metric in that versions of public reason liberalism that idealize citizens in order to reduce disagreement will be *pro tanto* inferior to those that do not. Thus reasonable pluralism can help to order conceptions of idealization. I set these metrics out in general terms here but their content will be specified throughout Part II as I employ them to uncover an alternative conception of public reason.

In applying the metrics, I will focus on different variables within the Public Justification Principle. Let us now decide the variables on which Part II will focus. To do so, recall again:

The Public Justification Principle: A coercive action C is justified iff it is a token of a rule(s) M that each and every member of the public P has (a) conclusive reason(s) R to endorse.

The Public Justification Principle contains variables C, M, P and R; these variables represent a range of values. In Chapter 1, I discussed some possible values, for instance by showing how conceptions

¹ I do not mean to exclude other metrics from playing a role in selecting a conception of public reason in other circumstances. But until the case is made for another metric and that the metric generates a problem for my argument, I will set these possibilities aside. I thank Chad Van Schoelandt for this point.

of idealization differ between, say, Rawls, Habermas and Gauthier or how to defend a convergence or consensus conception of reasons R. I have yet to defend specific values for C, M, P and R (coercive acts, scope of coercion, idealization and reasons). The first and second-personal metrics will aid in sifting through possible values of the variables. It is in this way that we can locate a plausible version of public reason liberalism that can effectively manage the conflict between social morality and individual ideal.

Part II is located attractive values for reasons R and members of the public P. The standard interpretation of these values involves adopting a consensus conception of reasons and a radical conception of idealization. The alternatives to these values are many, but they include versions of a convergence conception of reasons and more moderate conceptions of idealization. The first and second-person metrics will vindicate conceptions of public reasons that impose fewer costs on integrity and fewer restrictions on reasonable pluralism. For instance, values of P that significantly limit the scope of reasonable pluralism are *pro tanto* suspect on the second-personal metric. If a value for P idealizes citizens such that they always agree, then the value should be rejected if there is a less homogenizing alternative. Further, since integrity is significant, values for R and P must create significant social space for individual freedom. Values for R and P that restrict the expression of individual ideals are *pro tanto* inferior to those that do not.

I will not defend values for variables C and M. I suspect that whether a version of public reason liberalism can manage our conflict will be robust with respect to values for these variables. A brief examination of these variables will show why they need not concern us. First, consider coercive actions C. C refers to how finely we individuate units of coercion. We might hold that coercive *acts*, coercive *rules* or even coercive *institutions* are the objects of justification. The choice matters, but not as to whether the liberalism-faith conflict is manageable. The question of whether the tension is manageable depends upon whether conceptions of reasons and idealization can reconcile citizens by imposing a conception of social morality that preserves the rational freedom of citizens of faith. It is not clear how differing values for C would affect this outcome. For instance, three methods of filling out variable C populate the literature: constitutional essentials, coercive laws and units of coercion that lack causal dependence on other units.² The successful management of the liberalism-faith conflict does not depend on which of these options we select. More finely grained values will tend to be more hostile to coercion, which will more often restrict the integrity of citizens of faith to use coercion, but they will also restrict the coercion available to those with opposing values. Thus, it

² For the three positions, see Rawls 2005, p. 227. Quong 2004, pp. 233-250 and Gaus 2011, pp. 422-428.

is unclear how a choice of a value for C will affect the integrity of citizens of faith. If we can set aside seeking a value for C, we can do the same for M. M simply refers to the rules of which Cs are tokens and so the same considerations that permit us to set aside C seem to apply to M. To say that the management of the conflict does not hang on the values of M and C is just to say that nothing hangs on the values of M and C *over and above* what hangs on the values of the other variables.

The aim of Part II is Rousseau's: "To find a form of association that will defend and protect the person and goods of each associate with the full common force, and by means of which each, uniting with all, nevertheless *obey only himself and remain as free as before.*"³ But instead of finding a form of association, this dissertation takes a step back and tries to locate a *theory* of the form of association that will leave us as free as before. If we can find or construct such a theory, it will delineate the range of associations that can reconcile citizens of faith to liberal norms and institutions. It is in this way that the liberalism and faith conflict can be resolved.

³ Rousseau 1997, pp. 49-50. Emphasis mine.

Chapter 5: Conceptions of Public Reasons

I argued in Part I that the conflict between public reason liberalism and integralist liberalism is a case of the tension between first and second-person morality. In Part II, I construct a version of public reason liberalism that can resolve the tension. A viable strategy for resolving the tension requires reinterpreting the standard conception of reasons and idealization within the Public Justification Principle, which we understood as follows:

The Public Justification Principle: A coercive action C is justified iff each and every member of the public P has (a) conclusive reason(s) R to endorse C.¹

Again, Part II divides in half; the first half reinterprets reasons R and the second reinterprets the idealization of members of the public P. In this chapter, I will outline the standard conception of public reasons; Chapter 5 thereby sets the stage for Chapters 6 and 7 which criticize the standard conception and defend an alternative respectively.

Recall from Chapter 4 that a(n) (adequate) public reason is any intelligible reason that can enter into public justification. But Part II focuses exclusively on political theory. I take political theory to be that subset of moral theory focused on the proper use of coercion, so I will confine my discussion of public reasons to those public reasons relevant to the justification of coercion, not the broader set of reasons relevant to the justification of interference. We shall see that the literature in fact confines the set of public reasons *much* further. First, since many are suspicious of the use of integrity-based reasons in public life, they sometimes limit the set of public reasons to the set of second-person reasons. But they do so indirectly by confining the set of public reasons to those that are common among different citizens. Second, the standard conception of public reasons holds that they must be “shared” or “accessible”; other reasons are private and sectarian and thus inappropriate for use in public discourse and justification. Chapters 5-7 show why the set of public reasons should be extended beyond these narrow limits.

A conception of public reasons can be specified in accord with *requirements* concerning which subsets of reasons may be used to justify coercion, such as through a requirement of accessibility or shareability. In my view, the public reason literature implicitly or explicitly contains four such

¹ I drop the “M” variable to simplify our discussion.

requirements: (i) the intelligibility requirement, (ii) the accessibility requirement, (iii) the shareability requirement and (iv) the symmetry requirement. Chapter 5 analyzes and motivates these requirements; it also explains how some of them divide public reason and integralist liberals.

I: Intelligibility

I.i Exclusion and Restraint Revisited

Before introducing the various requirements on public reasons, we must revisit the distinction between principles of exclusion and principles of restraint. Recall that a principle of exclusion excludes a reason from entering into public justification; an excluded reason cannot serve as a basis of coercion. In contrast, a principle of restraint excludes a reason from being appropriately offered in public political discourse; restrained reasons should not be advanced or acted upon in the public sphere. The distinction raises the question of whether the standard requirements apply to exclusion, restraint or both. While the restraint-exclusion distinction is not recognized in the public reason literature, I believe that answer is clearly “both”. I will assume that the requirements exclude reasons, since the oft-cited problem with “non-public” or “private” reasons is that they cannot justify coercion. However, public reason liberals take the requirements to apply to restraint as well, since they habitually talk about the importance of right conduct in public dialogue.

The exclusion-restraint distinction determines how political philosophers should approach the matter of *sincerity*. Many public reason liberals think that only those reasons that can be sincerely offered as justifications are public.² It appears then that sincerity only applies to restraint, and is thus only relevant to the practice of reason-giving in the public sphere. Sincerity does not apply to exclusion because exclusion only concerns which reasons are suitable bases of coercion. The question of exclusion is therefore independent of whether citizens should offer reasons sincerely in the public sphere. Consequently, I will not include sincerity as one of the basic requirements on what makes a public political reason *bona fide*. Instead, sincerity must be motivated and grounded when the requirements are applied to dialogue. In my view, how and when citizens must be sincere is determined by principles of restraint and requirements on reasons already justified. I will therefore defer these matters to Chapter 6.

² For one prominent example of a principle of sincerity, see Schwartzman 2011, p. 12. For a related discussion, see Gaus 1996, p. 139.

I.ii Intelligibility

Public reason liberals typically maintain that public political reasons must be intelligible, evaluable, comprehensible, accessible, etc., to others. These distinct properties are often run together, but I shall group them into three degrees of stringency. The intelligibility requirement is the least stringent. As I argued in Chapter 4, by “intelligibility” I mean that public reasons must be ones that members of the public can see *as reasons* for those who advance (or rely upon) them, as opposed to mere utterances, expressions of emotions, irrational demands or other irrelevant considerations. Intelligibility is weak. It only requires that the public be able to see the reason as a reason according to the evaluative standards of the one who advances or relies upon it. Even if the reason in question would not count as a reason for a member of the public with another set of evaluative standards, the reason can still be seen as a reason for the person in question. The intelligibility requirement requires that those who reason from the standpoint of others recognize the considerations others offer *as reasons*. Without intelligibility, reasoning from the standpoint of others is impossible. Unless John can see Reba’s reasons *as reasons*, he can hardly take her perspective or consider the interrelations between her reasons and values.

Let us define intelligibility and the intelligibility requirement as follows:

Intelligibility: A’s reason X is intelligible to members of the public if and only if members of the public regard X as openly justified for A according to A’s evaluative standards.

Intelligibility Requirement: A’s reason X can justify coercing members of the public only if it is intelligible to them.

I take an “evaluative standard” to be whatever normative standard can be used to evaluate the contents of one’s rational commitments, values, beliefs, etc. and their inferential connections. I will assume for our purposes that reasonable pluralism applies to evaluative standards just as it applies to reasons. That is to say, reasonable people can disagree on their standards of inference and epistemic justification. For instance, some might regard theological discourse as intelligible by holding that one can reason well with respect to Biblical exegesis. Others may disagree, holding that most religions’ sacred texts are too indeterminate to effectively interpret. More abstractly, members of the public may differ as to what constitutes a good cognitive process. Many will worry that applying reasonable

pluralism to evaluative standards threatens to commit us to what Stephen Stich has called “normative cognitive pluralism” or the view that “various systems of cognitive processes that are very different from each other are equally good.”³ Gaus has argued that reasonable pluralism does not commit us to such an extreme position. While some cognitive standards differ, Gaus claims that psychological evidence shows that humans share a broad range of inferential rules.⁴ Gaus also argues that the fact that humans regularly and successfully reason together undermines the radical relativism of evaluative standards that Stich defends.⁵ Complex issues arise here that we must set aside. For now, I will take it for granted that the challenge of normative cognitive pluralism can be met.

The content of the intelligibility requirement will vary according to the selected conception of idealization, as the definition indicates. Recall that members of the public are idealized when the model of their reasons, cognitive capacities, inferential practices and information is improved to prevent cognitive errors from determining which reasons they possess. Thus, to determine whether A’s reason X is intelligible to members of the public, they must be idealized such that they have enough information and cognitive ability to evaluate whether reason X is openly justified to A. Political theorists will disagree about which level and what forms of idealization are appropriate, but some idealization is required if members of the public are to evaluate whether X is a reason for A according to A’s evaluative standards. At the least, they must have information sufficient to determine what A’s evaluative standards are, information not always available to actual members of the public. The intelligibility requirement also contains a second idealization condition because members of the public must ascertain which reasons are openly justified to A. The concept of open justification implicitly contains an idealizing element since it maintains that what is openly justified to A is justified in terms of chains of inference and additional information.

A critic will worry that the intelligibility requirement is vacuous because it will rarely restrict which considerations count as reasons. Two responses suggest themselves. First, the intelligibility requirement is entailed by the very idea of reasoning from the standpoint of others. If I cannot see your purported reasons as reasons for you even according to *your own* evaluative standards, then I cannot reason from your standpoint by definition. The point holds even if the intelligibility requirement is radically permissive. Second, intelligibility will rule out a number of purported reasons based on poor reasoning and bad information. In some cases members of the public may find that reasons advanced by A are in fact not reasons because they cannot see the putative reasons as

³ Stich 1990, p. 13.

⁴ Gaus 1996, pp. 47-48.

⁵ *Ibid.*, pp. 48-52.

justified to A on A's own terms. To take one example, suppose that John tells Reba he had an experience of Vishnu telling him to fight for racial equality and that therefore he has a reason to fight for racial equality. But suppose that Reba knows that John is a Christian. She will thus be warranted in assuming that John's reason is unintelligible because she knows that John is a Christian and so does not believe that Vishnu exists. Nonetheless, John may be aware of considerations and information that will not be included at the right level of idealization of members of the public. For instance, suppose we employ a moderate conception of idealization that only imputes information which can be gathered at relatively low cost. John's system may contain a complex series of inferential connections between his experience of Vishnu and his Christian beliefs that members of the public may not be aware of, even at the right level of idealization. For instance, John might have contact with the "St. Thomas" Christians of India who live in the province of Kerala. This community of Christians is nearly two thousand years old and often accommodates religious pluralism by integrating Buddhist and Hindu ideas into its theology. The St. Thomas Christians may have convinced John that experiences of Vishnu are compatible with Christianity.⁶ In this case, John may have a reason to believe in his experience of Vishnu because the reason is openly justified to him. However, his reason cannot enter into public justification because even at the right level of idealization it is unintelligible to members of the public, as they lack the information and processing capacity to discern the epistemic consistency of John's experience of Vishnu and his Christian beliefs. Thus, whether reason X is intelligible to members of the public does not always track whether reason X is a bona fide reason for A because A's rational commitments and the commitments members of the public can discern within A's system may differ.⁷

II: The Accessibility Requirement

The accessibility requirement is the next most stringent requirement in the public reason literature. Briefly, accessibility requires that a public reason be justified according to *common* evaluative standards. Rather than requiring that a public reason be justified according to the reasonable but distinct evaluative standards of members of the public, it requires that public reasons be justified

⁶ I personally have no idea whether this case plausibly holds of real St. Thomas Christians.

⁷ It is this sort of case that led me to distinguish between exemplary and adequate public reasons in Chapter 4. If Reba is aware that John is a St. Thomas Christian, then John's sighting of Vishnu will generate an exemplary reason because the reason will both *be* openly justified to John and Reba will *regard* the reason as openly justified as well. Therefore, what *is* openly justified and what is *taken to be* openly justified will match.

according to shared evaluative standards. Rawls argues that public justification requires that we share “guidelines of inquiry: principles of reasoning and rules of evidence in the light of which citizens are to decide whether substantive principles properly apply and to identify laws and policies that best satisfy them.”⁸ While accessibility permits *reasons* to differ it requires that they be evaluated as reasons according to evaluative standards that are shared. Thus, accessibility lies between intelligibility and shareability because intelligibility permits differing reasons and evaluative standards, whereas shareability permits neither.

Accessibility is perhaps the most common standard in the literature, with no less than eight available interpretations.⁹ For present purposes, we cannot review all eight criteria; but they are sufficiently similar that we need only analyze a subset: (i) in principle accessibility, (ii) criticizability, and (iii) independent confirmability.

II.i Three Standards of Accessibility

Let us begin with in principle accessibility:

- (1) A citizen’s reason R counts as a public justification for a given coercive law only if R is *publicly accessible* in principle.¹⁰

A reason is accessible in principle when persons can recognize the reasoning as justified from a common perspective. Christopher Eberle points out that this criterion requires not only that the “rationale be reliable; [one’s] compatriots must be able to see for themselves that the reason is sound.”¹¹ A “rationale”, I take it, is a set of reasons sufficient to justify a public political reason; “in principle accessibility” means that the rationale is accessible to persons at the appropriate level of idealization. Perhaps the reason will be accessible to all rational persons, or persons with full information. Eberle suggests that the rationale be publicly accessible to “competent and

⁸ Rawls 2005, p. 224.

⁹ Christopher Eberle has already outlined them in one place. See Eberle 2002, pp. 252-286. The accessibility requirement has been understood in terms of the following: intelligibility, accessibility, in principle accessibility, replicability, criticizability, dialogicality, independent confirmability, provability and proof of reliability. ‘Intelligibility’ here is distinct from the criterion I use.

¹⁰ Greenawalt and Perry use this criterion. See Perry 1993, Greenawalt 1988, p. 12, and Greenawalt 1995, pp. 151-164.

¹¹ Eberle 2002, p. 255.

levelheaded” observers.¹² But a clear conception of idealization is rarely provided when the standard is explained. In principle accessibility differs from intelligibility because it presumes that the aim of the reasoning is to provide a justification to all parties based on the same evaluative standards (even if the reason is rejected). Therefore, when Eberle says that our “compatriots must be able to see for themselves that the reason is sound” he means that the reason must be sound *for all*, not merely for the person offering it.¹³

Another common standard is criticizability, which we can define as follows:

- (2) A citizen’s reason R counts as a public justification for a given coercive law only if R is amenable to *external criticism*.¹⁴

Thomas Nagel defends a version of (2). He argues that we must not support coercive laws that lack a “common ground of justification.” Nagel writes that:

Public justification in a context of actual disagreement requires ... preparedness to submit one’s reasons to the criticisms of others, and to find that the exercise of a common critical rationality and consideration of evidence that can be shared will reveal that one is mistaken.¹⁵

Thus, we need a common evaluative standard to criticize proposals. This criterion requires that the rationale be criticizable by some shared standard of rationality and epistemic justification. A proposal based on a particular rationale may only be publicly justified if the rationale can be criticized by all others from their own perspective, but the external criticism criterion does not permit each individual to use her preferred evaluative standard. Instead, like in principle accessibility, a shared standard is required. The external criticism standard involves an idealization component given that the definition states that R “is amenable to” external criticism. The best way to understand “is amenable to” is to tie it to the proper level of idealization. Thus one who adopts this standard can say that the rationale R is accessible when it is amenable to external criticism or,

¹² For similar views, see Greenawalt 1988, p. 57.

¹³ I take it that Eberle does not mean that the reason is “sound” as an argument is sound, but rather that the reasoning possesses some epistemic credence.

¹⁴ Brower and Nagel defend versions of this criterion. See Brower 1994, p. 392 and Nagel 1987, p. 232.

¹⁵ Nagel 1987, p. 232.

equivalently, when it can be externally criticized at the right level of idealization. However, much like in principle accessibility, the proper level of idealization is often left underdetermined.

Nagel imposes a further restriction, claiming that “public justification requires ... an expectation that if others who do not share your belief are wrong, there is probably an explanation of that error which is not circular.”¹⁶ This second standard is motivated by a desire to avoid conflicts between “incompatible points of view.” Eberle therefore interprets Nagel as endorsing the following version of (2):

(2*) A citizen’s reason R counts as a public justification for a given coercive law only if (a) R is evaluable in light of a “common critical rationality” and (b) for any citizen who does not find R compelling, there exists a noncircular explanation as to why she does not find R compelling.¹⁷

Nagel notes that “conflicts of religious faith fail this test, and more empirical and many moral disagreements do not.”¹⁸ Non-circularity is required to bar disagreement from boiling down to one party simply disliking or rejecting the reasoning of another. Thus, the disagreement in question must result from a genuine difference in evidence and sound reasoning.

Let us now turn to independent confirmability:

(3) A citizen’s reason R counts as a public justification for a given coercive law only if R is amenable to *independent confirmation*.¹⁹

Public justification requires that the rationale for R can be somehow replicated by standards shared among persons other than the citizen who advances R. That is, others must be able to repeat the process of reasoning that led one person to R if R is to be a public reason. To illustrate, consider a putative religious experience: Saul’s conversion on the Road to Damascus. Only Saul sees Christ (although his company hears Christ’s words). Thus, it appears that Saul’s religious experience is not

¹⁶ Ibid.

¹⁷ There is a minor difference here between Nagel’s principle and Eberle’s formulation. Nagel requires that there be a *reasonable expectation* that disagreements will have non-circular explanations, whereas Eberle suggests that Nagel holds that there *in fact exists* such a noncircular explanation. I’m not sure this difference matters much for the purposes of the dissertation.

¹⁸ Nagel 1987, p. 270.

¹⁹ Greenawalt 1993, p. 649.

independently confirmable. Were Saul to justify a political policy by reference to his experience, his rationale would not be subject to independent confirmation. Independent confirmation is tied to external criticism, with the latter standard referring to the ability of others to evaluate reasons and the former to the ability to confirm them.

II.ii A Definition of Accessibility

The following definitions of accessibility and the accessibility requirement represent the standards above:

Accessibility: A's reason X is accessible to the public if and only if all members of the public regard X as openly justified to A according to common evaluative standards.

Accessibility Requirement: A's reason X can justify coercing members of the public only if X is accessible to all members of the public.

Three terms in the definition of accessibility need explanation. First, I employ the conception of open justification from Chapter 4. For accessibility, the conception of idealization will determine whether members of the public regard X as openly justified. So, like intelligibility, interpretations of accessibility will vary depending on the right level of idealization. The most important concept in the definition of accessibility, however, is the notion of "common evaluative standards." An "evaluative standard" is any normative standard that can evaluate the contents of a citizen's belief-value set as sound, unsound, consistent, inconsistent, and the like. Evaluative standards are "common" when they enjoy intersubjective recognition. Common evaluative standards are not necessarily "objective" standards.

Members of the public must regard reason X as openly justified *to A*. The accessibility requirement only requires citizens to advance reasons that they believe are justified according to recognized standards.²⁰ Nonetheless, the reasons may only be justified to the person who offers them. A reason can be justified *only for A* but in accord with common evaluative standards. For example, the scientific method is a common evaluative standard among scientists yet it might justify

²⁰ Here I speak of the requirement imposes constraints on what reasons citizens may advance, which implicitly assumes that accessibility is applied to restraint. The standard is usually taken to apply to both exclusion and restraint.

a scientific conclusion only to a sub-group of scientists given the ways in which they apply the standard to their particular data set.

II.iii Arguments for Accessibility

The case for accessibility is rarely developed in detail. One argument holds that giving others rationally accessible reasons expresses respect for their reasoning by attempting to reason from their standpoint. If we care about respecting others, we will offer them reasons that they can comprehend, that can appeal to them, or that they can assess, endorse or reject. When someone offers religious reasons on behalf of her favored policies, she appears to be uninterested in respecting those who cannot be expected to endorse her reasons.²¹ Christopher Eberle thinks these arguments imply that citizens should restrain themselves from using religious reasons out of a sense of reciprocity.²² As Nagel says, we are supposed to “present to others the basis of [our] own beliefs, so that once [we] have done so, *they have what [we] have*, and can arrive at a judgment on the same basis.”²³ But we do not share our religious views, so we cannot treat others fairly with religious reasons. It is thereby considered inappropriate to offer others reasons for supporting a coercive proposal that they cannot see as justified according to common evaluative standards.

For example, suppose that Roe proposes that abortion be banned on the grounds that after she had an abortion, she had a religious experience where God “convicted her heart” about her actions. Due to this experience, Roe concludes that abortion is wrong and that it should be banned. Now suppose that Roe is promoting a petition to place an anti-abortion proposition on her state ballot. She happens to knock on Gertrude’s door; Gertrude is pro-choice and believes that given their long history of oppression, women must aggressively and unapologetically defend their rights over their bodies, which includes the right to abortion. When Gertrude answers the door, Roe asks her to sign the petition. Gertrude, in a rage, exclaims, “Give me one good reason to ban abortion!” And Roe replies, “God says it’s wrong. He told me so after I had an abortion.” What is Gertrude to do in this situation except give Roe an incredulous stare? She cannot evaluate Roe’s reason as justified for herself. Not only does she not believe in God, but she senses that even if there is a God,

²¹ Notice that arguments for accessibility are often made with respect to restraint rather than exclusion; this fact simply illustrates the confusion in the literature.

²² Eberle 2002, pp., 109-151. We have already covered Gaus’s argument for the connection in another context. See Ch. 3, II.ii.

²³ Nagel 1987, p. 232. Emphasis mine.

He probably did not give Roe the message that abortion is wrong, despite Roe's conviction. In any event, she certainly does not take Roe's claim as evidence that abortion is wrong. Proponents of the accessibility requirement argue that Roe's actions are disrespectful because they do not make an attempt to offer Gertrude reasons she can be expected to accept. Roe's reasons are private, not public.

Religious reasons are often singled out as the paradigmatic private (qua inaccessible) reasons. Richard Rorty claims that religious argument is a "conversation-stopper."²⁴ Abner Greene sees religious grounds as a "secret box":

Imagine, for a moment, a group of citizens that has access to a box that contains evidence supporting a certain argument for a particular law. Suppose that group relies in the political process on the contents of that box but denies other citizens access to that box and its contents. We should exclude such shenanigans from politics because some citizens have access to the source of authority backing the law, while others are excluded from that source of authority ... Express reference to religious doctrine ... is the secret box model²⁵

Amy Gutmann and Dennis Thompson offer a similar line:

But the primary reason why [appeals to biblical authority in support of coercive laws] must be rejected as moral reasons [and thus as an appropriate basis for coercive laws] is that they close off any possibility of publicly assessing or interpreting the content of the claims put forward by the authority.²⁶

Similarly, Stanley Fish:

That is what neutrality means in the context of liberalism – a continual pushing away of orthodoxies, of beliefs not open to inquiry and correction – and that is why, in the name of neutrality, religious propositions must either be excluded from the marketplace or admitted only in ceremonial forms.²⁷

²⁴ Rorty 2009.

²⁵ Greene 1994, p. 659.

²⁶ Gutmann and Thompson 1996, p. 70.

²⁷ Fish 1996, p. 22.

If we do not offer others reasons that are accessible to them when we propose to coerce them, then we are not reasoning publicly and are thereby disrespectful.

II.iv Concerns about Accessibility

As argued in Ch. 2, accessibility is often taken to imply restraint on what can be said in a public forum. While principles of restraint vary in type and stringency, accessibility is used to justify most of them. If all public reasons must be accessible, then offering inaccessible reasons in public dialogue will be inappropriate or disrespectful. However, sometimes the accessibility requirement is applied to both exclusion *and* restraint (that is, to what can be justified or to what can be said). But accessibility can raise the integrity objection regardless of whether it is applied to exclusion or restraint. Integralist liberals reject restraint because it prevents them from speaking about and acting on reasons which are inaccessible to others, thereby violating their integrity. They will complain against exclusion that their reasons cannot legitimately influence the structure and impact of their institutions on their lives. Thus, if accessibility implies a significant integrity restriction via exclusion or restraint it will raise the objection with which this dissertation is concerned.

III: The Shareability Requirement

Shareability is the strongest of the three requirements, as it combines the requirement of shared evaluative standards with the requirement of shared reasons. Shareability also motivates a principle of restraint: we are only to offer others reasons they can share. Public reason liberals often argue that reasons should be shared but say little about what it means to share reasons. They presumably hold that reasons must be *shareable*, meaning that citizens will all have the reasons in question at the right level of idealization. Otherwise, requiring that reasons be actually shared would severely hamper political justification.²⁸ Due to this ambiguity, it is not clear what some public reason liberals mean when they speak of sharing reasons.

²⁸ Nonetheless, two recent articles defend an *actually* shared reasons standard. See Bohman and Richardson 2010, Schwartzman 2011.

III.i A Definition of Shareability

The shareability requirement, like intelligibility and accessibility, possesses two components—a requirement concerning the appropriate evaluative standards and a requirement concerning the range of permitted reasons. Shareability requires both shared evaluative standards and shared reasons. For our purposes, I will assume that if two citizens are able to share a reason this means that they are both rationally committed to regarding the reason as a reason. To put it another way, for A and B to be able to share reason X, X must be openly justified to both A and B. We may now define shareability and the shareability requirement as follows:

Shareability: A's reason X is shared with the public if and only if members of the public regard X as openly justified to the each member of the public, including A.

Shareability Requirement: A's reason X can justify coercing members of the public only if X is shared with them.

The concept of open justification remains familiar. I have simply extended the concept of open justification to apply to the entire public. A group shares reason X when X is openly justified to all of its members. However, a complexity arises. X can be openly justified to the public in two senses: (i) in terms of the shared beliefs and values of the public or (ii) in terms of the distinct beliefs and values of each member of the public. Shared reasons may be reasons that are openly justified according to the beliefs and values that citizens share, which is to say that there must be a sound chain of inference from these shared values and beliefs to the reason in question. Alternatively, shared reasons may be reasons that are openly justified to each individual given their distinct values and beliefs—there may be different sound paths to shared reasons. The set of shared reasons will be much larger on the latter interpretation. I do not believe that public reason literature is clear on which interpretation is preferred. So long as citizens use shared reasons, public reason liberals do not seem to be concerned how the reasons and values themselves are justified to each individual. Rawls seems to lean towards the less restrictive interpretation, as he suggests that the module of shared reasons modeled by the original position may be justified within the reasonable comprehensive

doctrines of a society in distinct ways.²⁹ For now, I will assume that the weaker interpretation is preferred.³⁰

III.ii Arguments for Shareability

There are three general arguments for shareability. The first argument is that shareability is implied by a commitment to respect for persons. The arguments from respect to shareability are similar to the analogous arguments for accessibility, so we can set them aside.

The second argument derives from the literature in deontological ethics, maintaining that a reason can only *count as* a reason if it is shareable. On this view, R is a bona fide reason for X only if it is a reason for Y. What it means to have a reason is that it be shareable. On this view, the shareability requirement is built into the very definition of a reason. The arguments for this claim tend to focus on the nature of reason claims and have been perhaps been most famously articulated by Thomas Nagel: “How can there be a reason not to twist someone’s arm which is not equally a reason to prevent his arm from being twisted by someone else?”³¹ For Nagel, it is hard to understand how there could be such reasons. He continues:

One would expect that reasons stemming from the interests of others would be neutral and not relative. How can a claim based on the interests of others not apply to those who may infringe it directly or intentionally in a way that it does not apply to those whose actions may damage that same interest just as much indirectly?³²

Nagel believes that human moral phenomenology reflects this sense that our reasons to act apply universally, at least those reasons that are interpersonal moral reasons.

Thomas Scanlon defends a similar view:

Since [Jane] accepts the judgment that, given G, she has reason to help her neighbor, Jane is also committed to the view that anyone else who stands in the relation described by G to

²⁹ Rawls 2005, p. 171.

³⁰ However, the second interpretation can raise the worries about the exemplary-adequate reason distinction raised in Chapter 4 as one member of the public A can regard X as shared with member of the public B for reasons that are not *actually* B’s reasons for endorsing the reason.

³¹ Nagel 1986, p. 178.

³² *Ibid.*

someone in need of help has reason to provide it. This is an instance of what I will call the universality of reason judgments.³³

Christine Korsgaard argues that the very idea of a private reason is incoherent, for reasons “are public in their very essence.”³⁴ My moral reasons by their nature apply to you as well. I will explore these arguments in more detail in Chapter 6. For now, if Nagel, Scanlon, and Korsgaard are correct, the shareability requirement has real bite; it excludes a massive number of putative reasons from public justification and the public sphere.³⁵

The third argument for shareability ties shareability to publicity and then points to the benefits of publicity. Recall Rawls’s three levels of publicity (Ch.3, III.iii). The first level of publicity “is achieved when society is effectively regulated by public principles of justice: citizens accept and know that others likewise accept those principles, and this in turn is publicly recognized.”³⁶ Everyone understands that a society’s basic structure is justified in terms of commonly accepted practices of inquiry and reasoning. The second level of publicity requires that citizens of a well-ordered society have shared beliefs “in the light of which first principles of justice themselves can be accepted, that is, the general beliefs about human nature and the way political social institutions generally work, and indeed all such beliefs relevant to political justice.”³⁷ The third level of publicity concerns what Rawls calls “the full justification of the public concept of justice as it would be presented in its own terms.”³⁸ This level must include all of the considerations we bring to the table when generating a conception of justice in the first place. The full justification must be “publicly known, or better, at least [known to be] publicly available.”³⁹ When a well-ordered society has all three types of publicity, Rawls claims that it meets “the full publicity condition.” Full publicity is an important ideal because it is appropriate to our status as “reasonable and rational citizens who are free and equal.”⁴⁰ In a society with full publicity, the exchange of reasons will render citizens “confident that this avowed reckoning itself will strengthen and not weaken public understanding.”⁴¹

³³ Scanlon 1998, p. 73.

³⁴ Korsgaard 1996, p. 135.

³⁵ In fact, if they are correct, the excluded entities aren’t reasons at all. I thank Chad van Schoelandt for this point.

³⁶ Rawls 2005, p. 66.

³⁷ *Ibid.*, p. 67.

³⁸ *Ibid.*

³⁹ *Ibid.* Again, Rawls allows for the weaker second condition because many may not want to expend the cognitive resources required to acquire the relevant knowledge.

⁴⁰ *Ibid.*

⁴¹ *Ibid.*, p. 68.

Micah Schwartzman has recently argued that publicity has great value for a number of reasons. First, it may be “necessary for democratic accountability” and it may well “enhance the quality of political decisions.”⁴² Publicity also makes justifications for coercion publicly available so they can be evaluated and examined.⁴³ To many, the advantage of a shareability requirement is that it promotes publicity. When citizens speak in shared terms, their arguments can be easily evaluated by their interlocutors because they share evaluative standards and reasons. Information is more easily shared when it is expressed in a common political language. Still, the connection between shareability and publicity is rarely developed. Some, like Schwartzman, seem to regard shareability as a constitutive part of publicity. For instance, Schwartzman claims that “justifications for political decisions should be ... based on shared or public reasons.”⁴⁴ Schwartzman seems to equate being a public reason with being a shared reason. But there is conceptual space between what is shareable and what is public since unshared reasons can be brought into the public sphere and recognized by others. But these issues must be delayed until Chapter 6, where they will be addressed in more detail.

III.iii Concerns about Shareability

Critics of public reason insist that the shareability requirement imposes excessive burdens on citizens. Much like accessibility, shareability is taken to imply principles of restraint that impose integrity costs on many, including citizens of faith. If reasons offered in public must be accessible, then citizens are merely required to offer reasons that they believe can be justified in accord with common evaluative standards. But if reasons offered in public must be shareable, then citizens are required to only offer reasons that they believe others are rationally committed to. Such a judgment is much harder to make as it requires a significant amount of information. To offer a shareable reason in good conscience, a citizen must know a significant amount about her interlocutor. Further, the reasons she may offer will be substantially restricted. Due to reasonable pluralism, citizens may not share many important reasons. Consequently, restraint may impose substantial burdens on citizens. With accessibility, unshared reasons may be advanced so long as they can be impartially justified. If the accessibility requirement excessively burdens citizens of faith, the shareability requirement will be much worse. We can expect therefore that any integrity-based case against accessibility will probably apply more powerfully to shareability.

⁴² Schwartzman 2011, pp. 7-8.

⁴³ *Ibid.*, p. 9.

⁴⁴ *Ibid.*, p. 8.

III.iv A Diagram

We have now covered three of the central requirements on public reasons within the literature. I find it illustrative to group them together in a diagram:

		Evaluative Standards	
		Unshared	Shared
Reasons	Unshared	Intelligibility	Accessibility
	Shared	X	Shareability

The diagram contains two axes, one that covers reasons and the other that covers evaluative standards. A requirement can mandate that reasons be shared and/or that evaluative standards be shared. Requirements can also permit reasons to be unshared and evaluative standards to be unshared. To review, intelligibility requires that neither evaluative standards nor reasons need be shared, whereas accessibility requires that evaluative standards be shared, but not reasons. Finally, shareability requires both shared reasons and shared evaluative standards. There is logical space for a fourth requirement, but I leave it unanalyzed as it seems difficult to motivate.

IV: The Symmetry Requirement

The symmetry requirement is distinct from the intelligibility, accessibility and shareability requirements. It does not restrict the range of permissible reasons based on whether they possess some epistemic property like shareability. Instead, symmetry requires that public reasons be used symmetrically across a number of contexts. Specifically, the symmetry requirement holds that the

standards that apply to reasons that propose coercive action are equivalent to reasons that reject or defeat coercive action. We may therefore define symmetry as follows:

Symmetry Requirement: The same evaluative standards must hold for reasons to propose coercive action C as the reasons to reject coercive action C.

Delineate two classes of reasons, reasons to propose coercive actions (P) and reasons to reject or defeat coercive actions (D). Suppose that a shareability requirement is justified for P. If symmetry holds, then shareability must apply to D as well. In fact, if symmetry holds, any standard of reasons must apply symmetrically, that is, any version of the intelligibility, accessibility or shareability requirements must hold across sets of reasons P and D. If we deny symmetry, one standard could apply to P and another to D. For instance, if symmetry is rejected, set P may have to be shareable, whereas set D need only need be intelligible.

I am unaware of any place in the public reason literature where the symmetry requirement is defended or even articulated, but it appears to be implicit in the standard conception of public reasons. One reason to think that symmetry is implicit is that public reason liberals do not recognize that an argument for restrictions on P might not be an argument for restrictions on D. For this reason, I can only speculate on the motivation for symmetry. If asked, public reason liberals may simply argue that the considerations that motivate a restriction on reasons to public political reasons to propose coercion will motivate the restriction on public political reasons to reject (or vice versa). By assuming shareability or accessibility, they might also assume that the common pool of justificatory reasons is small enough to where there will be little or no relevant difference between reasons to propose and reasons to reject coercion.

V: The Standard Conception of Public Reasons

This chapter has argued that the standard conception of reasons within public reason liberalism is composed of four requirements. I will appeal to these definitions throughout the dissertation:

Intelligibility Requirement: A's reason X can justify coercing all members of the public only if it is intelligible to them. A's reason X is intelligible to members of the public if and only if the public regard X as openly justified for A according to A's evaluative standards.

Accessibility Requirement: A's reason X can justify coercing all members of the public only if X is accessible to them. A's reason X is accessible to the public if and only if members of the public regard X as openly justified to A according to common evaluative standards.

Shareability Requirement: A's reason X can justify coercing members of the public only if X is shared with them. A's reason X is shared with the public if and only if X is openly justified to each member of the public, including A.

Symmetry Requirement: The same evaluative standards must hold for reasons to propose coercive action C as the reasons to reject coercive action C.

Public reason liberals commonly hold all four requirements implicitly or explicitly, though some reject shareability.⁴⁵ The accessibility requirement is widely held in some form or another. And the intelligibility requirement is often mixed with accessibility. The symmetry requirement, again, appears to be implicit in the literature. From here forward, I will call a conception of reasons that combines all four requirements the *strong symmetric consensus* conception of reasons. It is "strong" because it requires shareability, and it is symmetric because it endorses symmetry. A *weak symmetric consensus* conception of reasons rejects shareability for accessibility. A conception of reasons that affirms only intelligibility I shall call the *convergence* conception of reasons. In Chapter 6, I will argue against symmetry, shareability and accessibility. In Chapter 7, I defend convergence.

⁴⁵ Robert Audi rejects shareability, to give one prominent example. See Audi 2000, Ch.1 for Audi's "plural paths" to liberalism. It must be noted here that Audi does not fit squarely within the public reason liberal tradition.

Chapter 6: Against the Standard Conception of Public Reasons

The standard conception of public reasons includes the accessibility, intelligibility and symmetry requirements. Many versions of the standard conception include the shareability requirement as well. In this chapter, I argue that we should reject accessibility, shareability and symmetry. The defense of an alternative conception of public reasons is left to Chapter 7.

This chapter's arguments follow the strategy for assessing conceptions of public justification developed in Part I. Recall that the Public Justification Principle does not entail any particular conception of public reasons. Rather, it delineates a set of eligible interpretations. These interpretations can be ranked according to first and second-personal criteria.¹ Conceptions of public reasons that impose relatively few costs to integrity or that better express the foundational values of public reason are superior to those that impose relatively more. I will argue below (I) that requirements that fail to achieve their aims are inferior to those that do not. We will see that the aims for supporting a requirement are often not achieved by adopting the requirement. Requirements failing to achieve the aims for which they were adopted should be rejected. The strategy of the chapter is to show that conceptions of public reasons containing the shareability, accessibility and symmetry requirements are inferior to those that lack them because the three requirements either fail to achieve their aims or impose more first and second-personal costs than alternatives.

The bulk of the chapter is built around the argument against accessibility. The argument, briefly, is that *the accessibility requirement either fails to achieve its aims or is excessively restrictive on the first and second-personal metrics*. I defend the thesis by showing first that plausible versions of the accessibility requirement allow appeal to practically any reason, including many controversial religious reasons. Most plausible interpretations of the accessibility requirement are committed to regarding even reasons deriving from natural theology and religious testimony as accessible. If the accessibility requirement cannot exclude the *paradigmatic* private reasons—religious reasons—then it fails to achieve its aim of separating seeming public from private reasons. I then argue that reformulating

¹ Throughout the dissertation, I understand A as ranked above B whenever the balance of first and second-personal considerations favor A over B. Ranking need not imply an ordinal or cardinal scale.

the accessibility requirement to exclude these reasons makes it implausibly restrictive with regard to the second and first-personal metrics.

Christopher Eberle has developed a similar dilemma for accessibility requirements.² Eberle claims that mystical experiences can provide reasons that satisfy the accessibility requirement, however construed. So why repeat the argument? Eberle's criticism of public reason liberalism relies on Reformed epistemology.³ Reformed epistemologists believe that mystical experiences provide reason to believe in supernatural beings, their presence and even their testimony. Public reason liberals will tend to resist this view. Eberle's argument is thereby weakened because it presupposes that mystical experiences provide sound reasons for belief in the supernatural. One might worry that Eberle's argument can be reformulated for less tendentious epistemological positions, but this has yet to be shown. Consequently, many will not find his argument appealing. To avoid this problem, I discuss forms of religious reasoning whose soundness relies on fewer substantive epistemic commitments, i.e. reasoning based upon natural theology and religious testimony.

Accessibility is less restrictive than shareability on the first and second-personal metrics. Accordingly, if accessibility should be rejected, shareability should as well. If accessibility and shareability are defeated, the case for symmetry will be eliminated. For without accessibility and shareability, there is nothing for the symmetry requirement to do. With the requirements analyzed and rejected, we will have the resources to give a satisfying analysis of the heretofore unexamined *sincerity* requirement on public reasons. I will show that sincerity reduces to the combination of the foregoing requirements with a principle of restraint. The machinery of this chapter and the last can then show that sincerity is a red herring.

Since the requirements have already been analyzed (Ch.5, II-IV) and a conception of religion, faith and religious reasons already presented (Ch.2, I.ii), I shall first consider what counts as a good reason to reject accessibility, shareability or symmetry. In Section II, I will advance a challenge to accessibility based on natural theological arguments. In Section III, I will advance the challenge of religious testimony. Section IV attempts to save the accessibility requirement by modifying the level of idealization, whereas Section V assesses the shareability requirement as a potential alternative. Section VI marshals the foregoing to reject the symmetry requirement. Section VII analyzes the sincerity requirement. Section VIII concludes.

² Eberle prefers to reject public reason liberalism rather than accessibility. See Eberle 2002, pp. 234-293.

³ Two canonical works in the Reformed Epistemological tradition are Alston 1991 and Plantinga 2000.

I: Good Reasons to Reject Requirements

I have argued above (Ch.5, I) that only the intelligibility requirement is manifestly entailed by a commitment to public justification. The other requirements must be inferred from additional premises. As we have seen (Ch.4, II.i, II.ii; Ch.5, II.iii, III.ii, IV), defenses of these additional requirements are inconclusive because they are not entailed by the Public Justification Principle. Consequently, a commitment to public justification admits of various *conceptions* of public reasons, some including accessibility, shareability and symmetry and some not. Further arguments must decide which conception is best.

The first and second-personal metrics can decide which conceptions of public reasons are superior to others (Ch.4, V.iii). To review, the first-personal criterion concerns integrity-related costs; so, a conception of public reasons is superior to another conception on this criterion if it places fewer burdens upon integrity. Respect for integrity requires that fewer burdens be placed on integrity where possible, at least when the constraints of other criteria are held constant. The preservation of integrity and the ability to act on first-personal reasons are among the main reasons to participate in liberal social morality. Accordingly, a commitment to public justification already presumes respect for integrity. Thus, if a conception of public reasons places fewer burdens on integrity, that is a strong mark in its favor. The second-personal criterion holds that a conception of public reasons is superior to others when it better expresses the core values of public reason, such as the recognition of reasonable pluralism. When a conception of public reasons is inferior to other eligible conceptions on the second-person criterion, it should also be rejected.

Moreover, a conception of public reasons should be rejected if it *fails to achieve its aims* or at least *adds nothing to achieving its aims*. As such, if two conceptions of public reasons, A and B, count a similar set of reasons as public, but the point of A is to be somewhat restrictive vis-à-vis B, then A is inferior to B. For example, suppose we compare two conceptions of public reasons: (i) a conception that contains a shareability requirement and an accessibility requirement both with radical degrees of idealization and (ii) a conception that contains only an accessibility requirement with a radical degree of idealization. Since both conceptions invoke radical idealization, they may generate similar sets of reasons since many models of radical idealization idealize until all members of the public reason identically. The shareability requirement would thereby fail to mark out a distinctive subset of shared reasons as all reasons would be shared. The shareability requirement's aim is to restrict justificatory reasons to a shared set in order to preserve some sense of a shared point of view in politics. By

failing to produce any further restriction beyond accessibility, it has at least added nothing to its aim. Hence, conception (ii) is preferred to conception (i).

My strategy for defeating requirements needs further clarification. Suppose that shareability is indeed excessively restrictive of integrity. It does not follow from this that *every* conception of public reasons with shareability is inferior to *every* conception of public reasons without it. Rather, it means that for any two conceptions that are similar in every respect *save shareability*, the conception that lacks shareability is superior to the conception that contains it. Thus, if shareability is more restrictive of integrity, then for any conception that contains shareability there exists a conception with the same requirements *save shareability* that is superior to it. Accordingly, any time a conception of public reasons is offered that contains shareability we may always point to a conception that is preferred to it on first-personal grounds even if some conceptions of public reasons that contain shareability are better than those without it. To put it another way: it does not matter if a conception of public reasons containing only accessibility at a very high degree of idealization is inferior on the first-personal metric to a conception containing accessibility and shareability at a more moderate degree of idealization. There will still always be a conception of public reasons containing accessibility alone that is superior with respect to the first-personal metric to any conception with accessibility and shareability. If conception A contains shareability at degree of idealization X and accessibility at degree of idealization Y, and conception B contains only accessibility at degree of idealization Y, then conception B is superior to conception A, *for any X restrictive of integrity*.⁴ The same reasoning will hold for vindicating other conceptions of public reasons on the first-personal metric *and* the second-personal metric.

I shall now argue that the accessibility requirement either fails to achieve its aims or is excessively restrictive on the first and second-personal metrics. Depending on how it is interpreted, accessibility should be rejected because it (i) fails to achieve its aim, (ii) incurs high first-personal costs, or (iii) incurs high second-personal costs. As we have seen (Ch.2, II.ii; Ch.5, II.iii), standard interpretations of accessibility regard only secular reasons as accessible. These interpretations are thereby thought to entail principles of restraint which, in turn, place burdens on integrity. If so, conceptions of public reason that contain accessibility should be ranked below those without it on integrity-related grounds. However, I claim that accessibility must treat secular and religious reasons

⁴ Furthermore, the reasoning will hold for all the requirements even when the set of requirements is highly individuated by level of idealization. Thus, if shareability at degree of idealization X is, say, redundant, then there is always some conception of public reasons that lacks shareability at degree of idealization X that is superior to it.

similarly, that is, it must either count both types of reasons as accessible or neither. The accessibility requirement must therefore either place burdens on the use of religious *and* secular reasons or few on either. If the burdens are low, then accessibility seems to lose its rationale of marking out a distinctive class of public reasons. In other words, we lose purchase on the commonsense understanding of when a reason is accessible. If the burdens are high, however, first and second-personal considerations show that conceptions of public reasons with accessibility are inferior to conceptions without it. Most of the argument defends the claim that accessibility fails to achieve its aims. The remainder shows that modifying its ordinary level of idealization makes accessibility implausibly restrictive on the first and second-personal metrics. Thus, by the end of Section IV, we shall have *all three grounds* to reject accessibility.

I have heretofore claimed that I will show that accessibility fails to achieve its aims on most plausible interpretations of it. What does this mean? Given the foregoing (Ch.5, II.iii), I take the aim of accessibility to be somehow “separating the public wheat from the private chaff.”⁵ In other words, accessibility is supposed to capture some intuitive distinction between reasons that we can all evaluate as justified and some set of reasons that we cannot. When Abner Greene calls religious reasons a “secret box” he has this intuitive distinction in mind. Somehow religious reasons in particular are inaccessible to those who do not share those reasons. It is hard to give much substance to this intuition, as I believe it evaporates upon closer examination. But the general claim is that there is some sense to be made over and above the work done by intelligibility of a reason being inaccessible, particularly of a sectarian or religious reason being inaccessible. My aim in this chapter is to demonstrate that this intuitive distinction cannot be defended in the face of sustained criticism.

Despite discussing these arguments against accessibility (Ch.5, II.iii), we have already seen that the arguments fail to establish that accessibility is entailed by a commitment to reasoning publicly (Ch.2, II.i). I also noted in Chapter 5 that arguments for accessibility are usually generated by drawing analogies with civil public discourse. Mainstream public reason liberals argue that it is disrespectful for citizens to offer others considerations that they do not share or cannot access. But we saw throughout Chapter 2 that such arguments run together accessibility, exclusion and restraint; no argument for accessibility apart from arguments for restraint is offered. When mainstream public reason liberals begin the argument for accessibility with claims about dialogue, they are moving “upstream” from purported facts about what is appropriate to say in public, to what reasons must

⁵ Eberle 2002, p. 14.

be excluded, and then to whether public reasons are by their nature accessible. In contrast, I have pursued a “downstream” approach, arguing that the case for restraint derives from the case for exclusion, which in turn derives from the case for accessibility or shareability. Arguments from accessibility and shareability must therefore derive from a commitment to public justification. Recall (Ch.5, II.i) the three entailments I argued must be established to make the case for principles of restraint:

- (1) Public Reason Liberalism → Accessibility/Shareability Requirement
- (2) Accessibility/Shareability Requirement → Principles of Exclusion
- (3) Principles of Exclusion → Principles of Restraint

Since practically *all* of the arguments for accessibility are generated from intuitions about proper restraint, none of these arguments address whether public reasons are by nature accessible. Further, mainstream public reason liberals do not offer any argument of the sort I have developed in the dissertation. In other words, they do not admit that the Public Justification Principle is inherently ambiguous between different conceptions. Accordingly, they do not argue that conceptions of public reason with accessibility are superior to those without it because they do not acknowledge the conceptual space required to see that such an argument is necessary. Public reason liberals come closest to arguing for accessibility when they argue for *sincerity*, but I argue below (VII) that sincerity itself is derived by combining accessibility or shareability with a principle of restraint. Since the case for conceptions of public reason with accessibility is presently unavailable, let’s turn to the first argument for the claim that accessibility cannot meet its aims, the challenge of natural theology.

II: The Challenge of Natural Theology

I shall now argue that normal interpretations of accessibility must count natural theological reasons as accessible. Showing the admissibility of such reasons is the first step towards showing that accessibility cannot meet its aims.

Natural theology is the attempt to discern evidence for the existence or activity of the supernatural through natural reason.⁶ Branches of natural theology pursue a priori argumentation for

⁶ George Howard Joyce, S.J. describes natural theology as “that branch of philosophy which investigates what human reason unaided by revelation can tell us concerning God.” Joyce 1922, p. 1.

the existence of God or defenses of theological claims concerning the nature of God or God's will. Some strands of natural theology concern whether one can have good reason to believe in Divine revelation or the relation God bears to the human soul.⁷ Further, natural theologians claim to be able to compete with secular philosophers in the realm of pure reason. Their claim is not implausible. For instance, Robert Audi, among the most prominent mainstream public reason liberals has recently modified his view to count some natural theological claims as permissible public reasons.⁸ On his view, the reasoning of natural theologians is evidentially on par with secular reasoning.⁹ And many of the great historical philosophers like Thomas Aquinas make some theological arguments that can be evaluated, accepted or rejected on rational grounds alone, i.e. with no appeal to revelation (or any other evidence in a "secret box").¹⁰ A general characterization of a political argument based on natural theological claims, then, is one that either entirely or partly depends on a natural theological claim for its soundness. Thus, a natural theological argument for teaching intelligent design in the classroom might hold that natural reason can demonstrate the need for an intelligent designer in nature with the powers traditionally ascribed to the Judeo-Christian God.

If natural theological arguments appeal to natural reason alone, their arguments can be evaluated by common evaluative standards and assigned epistemic justification on that basis. Consequently, such arguments will count as accessible, posing a challenge to accessibility's putative stricture against religious reasons. In evaluating the accessibility of natural theological arguments, I shall therefore hold them as justified when reasonable, well-informed citizens can endorse them after a degree of reasoning appropriate for citizens engage in public political life. Public reason liberals should not require that a citizen reason extensively about a particular argument, gathering all the information he can and processing it to the absolute best of his ability. Instead, the appropriate epistemic standard for public argument is one that we can reasonably expect our fellow citizens to live up to.

It matters that natural theological arguments are accessible. Given that natural theology is usually devoted to demonstrating the existence of the God, it may have important political

⁷ See Swinburne 1997 and Swinburne 2007.

⁸ It is probably more accurate to describe Audi as a fellow traveler of the public reason tradition, not as a member.

⁹ Audi stresses that natural reason is not essentially religious but can establish religious conclusions, such as that God exists. He is skeptical that it can be used to establish moral and political conclusions without appealing to revelation. He also stresses that his notion of natural reason does not depend on the metaphysical proposition that we could comprehend or discover natural reasons without divine aid. Audi 2009, p. 165.

¹⁰ I do not mean to imply that faith is irrational; I merely mean that natural theology does not rely on supernatural testimony, i.e. revelation. Audi expresses a similar position, arguing that separating natural reason and theology is not meant to "imply any disrespect for theology." Ibid., p. 167.

ramifications. Audi worries that if natural reason is taken to be capable of establishing theism, “then the way is open to hold that governmental establishment of at least a generic theism is justifiable independently of any particular religion.”¹¹ If one has good reason to believe God exists, then she may think that she can discern God’s will and conform her behavior—and potentially the behavior of others—to that will.¹² Consider an example. Catholic theologians often claim that there is reason to suppose that God provides a fetus with a soul at conception. Presence of the soul creates personhood in the fetus. Hence, the fetus is a person and must not be killed. Suppose, then, that a traditional Catholic wishes to defend her vote against permitting abortions. Her argument might go as follows:

- 1) The existence of God can be rationally demonstrated.¹³
- 2) God gives each human body a soul that can survive death and provides a human life with intrinsic worth.¹⁴
- 3) The least arbitrary candidate for the union of soul and body is the first presence of a unique biological potentiality, i.e. conception.
- 4) Thus, persons with intrinsic moral worth exist at conception.
- 5) All persons, therefore, fetuses included, must not be destroyed.

This argument is based on inferences that purport to require no appeal to revelation. While the above argument is not valid, it represents the reasoning of an ordinary, but reasonably well-informed, citizen. We can therefore still regard the argument as sound by proper standards of public argument. To begin, consider whether one could justifiably believe the above argument based on common evaluative standards, despite the fact that the argument may be mistaken. I submit that the answer is yes. If so, this should be illustrative. If the defenses of the premises are straightforward and based on good reasoning, then properly idealized members of liberal societies will be able to access the argument. Thus, religious reasons will satisfy the accessibility requirement at most levels of idealization even if they cannot ultimately justify coercion.

¹¹ Ibid., p. 166.

¹² Audi argues that natural theology cannot establish any “specific moral and political standards” without appealing to a religious tradition. Ibid., p. 168.

¹³ The Catholic Church has maintained for centuries that the existence of God can be demonstrated by an appeal to natural reason. Boniface VIII 1302. One classic argument can be found in Aquinas [1268] 1975, pp. 139-141

¹⁴ Several theistic philosophers have maintained that the ensoulment hypothesis is one of the best ways to explain the “fact” that we have souls. Swinburne 1997, pp. 174-199.

Consider the first premise: the existence of God can be rationally demonstrated. This is an ordinary claim. Many affirm it, many deny it, and the rest have no settled view. But most reasonable persons acknowledge that positive arguments for the position cannot be immediately dismissed. There are many reputable arguments for God's existence, even if they ultimately fail. For instance, versions of the cosmological and teleological arguments are widespread, even in conversation about religious matters with ordinary citizens. Both arguments are routinely analyzed and evaluated by people with distinct views (some of whom are analytic philosophers). If those involved in the discussion acknowledge that reasonable people can disagree with them, then they should regard many of those who accept the arguments as justified in doing so. This is not to say that the mere fact that a reasonable person could believe P shows that P is accessible, though; instead, it is only to claim that the recognition of reasonable disagreement in this case is evidence that some people justifiably affirm the argument and thus that the reasons derived from the argument could be sound to those who accept the argument.¹⁵

Consider the second premise: God provides human bodies with souls that give them intrinsic moral worth. Billions of humans believe in God and many more believe in at least one god; furthermore, some psychological evidence suggests that theistic belief comes naturally to us.¹⁶ The vast majority of humans also believe in some kind of soul.¹⁷ Arguably at least *some* reflective people have reasoned carefully enough to rationally believe that God exists, creates souls, and somehow attaches them to human bodies and consider the matter enough to acquire epistemic justification for their views. These views may be *false*, and they may not be well-supported by the most expansive and clear-headed understanding of the evidence, but neither truth nor exhaustive reasoning are the appropriate standard of accessibility. Instead, public reason liberals seek a degree of epistemic justification appropriate to ordinary political argument (see Ch.4, II.i). This is not to say that all widely-held beliefs qualify but rather that epistemic justification can be achieved for beliefs and reasons appropriate for public argument through the amount of reasoning that we can appropriately demand of our fellow citizens.¹⁸ Consequently, such views need not fail to have some positive epistemic status. Surely a large proportion of those individuals have epistemic justification for their views.

¹⁵ I thank Chad van Schoelandt for this point.

¹⁶ See Barrett 2004 for an attempt to show that theistic belief is cognitively natural for humans.

¹⁷ See Bloom 2005 for an argument that belief in the soul comes naturally to humans.

¹⁸ I explain this standard at length below (Ch.10, I).

The third premise holds that God probably attaches souls to bodies at conception. While the arguments for this position are complex, many pro-life theorists have defended conception as the least arbitrary point for God to attach souls to bodies. It is not hard to see that this is one reasonable view for reflective individuals, even if there are others. At the moment of conception a unity is created and has a certain biological potentiality. The ensoulment point may occur elsewhere in development, but conception is not an unreasonable starting point. The fourth premise only relies on the view that persons have dignity and are inviolable, a view widely affirmed by reflective reasoners, religious or secular. Many believe that possessing a soul is essential to personhood, and so the presence of a soul entails the presence of a person. If fetuses are persons, then it is easy to see why they should not be killed.¹⁹

Many *unidealized* citizens can evaluate the argument and come to see it as justified for others, including reflective non-religious citizens. Granted, the argument is bound to be controversial, but a rationale is not disqualified merely because it is contentious. What matters is that each premise in the argument can be evaluated and assigned positive epistemic status. So: is the argument we have discussed accessible at reasonable idealization-values? The answer seems to be yes; the defenses of the premises appear to be those that a reflective and rational citizen could endorse as evidenced by the fact that many reflective and rational citizens in fact *do* endorse them after thoughtful consideration. Consequently, the natural theological argument against abortion meets the accessibility requirement on many reasonable interpretations, despite its religious content. The point can be expanded to other natural theological arguments. Many argue against the moral permissibility of suicide on natural theological grounds.²⁰ A reasonable citizen may hold that God is the sole moral authority over life and death and conclude that suicide is impermissible. This argument is no less controversial than the argument against abortion, but its premises can presumably achieve the same level of epistemic credence. Both arguments may be *inconclusive*. But most arguments advanced in the public square are inconclusive.²¹ We cannot rule out candidates for public justification on such grounds.²²

¹⁹ Of course, philosophers like Judith Jarvis Thomson argue that even if fetuses are persons, that abortion should still be permitted. But I assume that many reasonable persons will reject Thomson's argument. See Thomson 1971.

²⁰ John Locke makes one such argument, that an individual "has not liberty to destroy himself" because God has not authorized him to take such an action. Locke 2003, p. 102.

²¹ Gerald Gaus argues this point persuasively. See Gaus 1996, pp. 152-3.

²² The arguments might also be *defeated* within the belief systems of those advancing the argument. If all religious rationales were so defeated, then they would not be admissible into public reason. I presume that not all of them will be, given the public reason liberal commitment to reasonable pluralism.

From the argument above, it seems that natural theological arguments not only satisfy the accessibility requirement but can be used to establish substantive moral and political principles if we only require of citizens a standard of rationality and information appropriate for public discourse. Reviewing other natural theological arguments is unnecessary. From the foregoing we can infer that some natural theological reasons satisfy the accessibility requirement. Let us move to the challenge of religious testimony.

III: The Challenge of Religious Testimony

A stronger reason to suppose that the accessibility requirement cannot satisfy its own aims is that arguments from religious testimony count as accessible despite the fact that they are the *paradigmatic* private reasons. If reasons deriving from religious testimony are accessible, then it is no longer clear whether accessibility does the work it is intended to do.

Religious testimony can be defined as any statement or utterance concerning the action of or communication with supernatural agents. Sacred texts that record testimonies count along with testimony by authorities who purport to have contact with divine beings. Examples of religious testimony include, among others, the ex-cathedra infallible pronouncements of the Papacy and Muhammad's link with the archangel Gabriel, as do the Torah and the Bible.

Consider Teresa, a Christian who deems homosexual practices morally impermissible. Suppose that the basis of this belief is her reading of Romans 1. In the passage, the Apostle Paul testifies that the reason God destroyed Sodom and Gomorrah was due in part to homosexual behavior practiced and tolerated there.²³ As a result Teresa argues like so:

- (i) The Bible is the central communication of God to humanity.
- (ii) The Bible is therefore infallible.
- (iii) The Bible teaches that homosexual practices are morally impermissible.
- (iv) Therefore, homosexual practices are morally impermissible.

²³ This is a matter of some controversy among Biblical scholars. See Boswell 1981 and Gagnon 2004.

Due to this argument, Teresa decides to support a ban on homosexual marriage and votes against the repeal of sodomy laws in her state. Does Teresa thereby rely on a religious rationale that fails to meet the accessibility requirement?²⁴

Teresa does not appear to share the epistemic situation of the citizen employing reasons of natural theology. The first premise does not have as long a history of philosophical argument on its behalf, although it would be a mistake to say that it has never been rationally defended. Nonetheless, many will view premise (i) as a paradigm violation of the accessibility requirement. All non-Christian citizens will reject the proposition and many will regard it as unreasonable. It also appears that the standard reasons for believing premise (i) are ones that are not accessible to those who reasonably disagree. Sometimes people believe the proposition for no reason at all or by testimony alone.

Premise (i) can satisfy the accessibility requirement at the proper idealization value in two ways. First, it can be attached to reasons of natural theology. Teresa's fellow citizens might find her rationale for premise (i) accessible if she could situate it within an argument for God's existence and a further argument that the Bible is reliable testimony of God's will. Some reasonable persons have defended arguments for God's existence.²⁵ Further, philosophers of religion often defend the view that God's existence entails His goodness.²⁶ Several theologians and philosophers across history have argued that a good God would communicate with and aid His creatures.²⁷ These arguments are sometimes accompanied by arguments that the best candidate for revelation is the Bible.²⁸ All of these arguments proceed by deductive and inductive inferences, and the chains of reasoning contained in these works are not clearly unjustified. Since premise (iii) is a claim about how to interpret Bible it appears accessible as well.²⁹ Evidence for this includes the fact that many non-Christians have engaged in the dispute.³⁰ Therefore, even many non-Christians can evaluate the reasons and assign them positive epistemic status. Premise (iv) flows naturally from premises (ii) and (iii). If God is good, has revealed His will to us in the Bible and the Bible says that homosexuality is wrong, then homosexuality is probably wrong. A good, honest God who tells us that homosexuality is wrong would presumably know whether it was wrong and tell us the truth about it. The argument

²⁴ I should state here that merely because these reasons are accessible does not mean that, on my account, banning sodomy could be publicly justified.

²⁵ Many philosophers do to this day. See Moreland 2009, Plantinga 1990, and Swinburne 2004 for some recent attempts.

²⁶ Again, see Aquinas [1268] 1975, pp. 139-141. For a sympathetic analysis of the argument, see Kretzmann 1997.

²⁷ Swinburne 2007, pp. 79-106.

²⁸ *Ibid.*, pp. 239-288.

²⁹ See. ft. 24.

³⁰ For a study of the argument, see Mize 1996.

is not deductive as it stands, but it meets the justificatory standard required to be permissible in public justification.

Reasons derived from religious testimony like premise (i) are also accessible because they are analogous with reasons derived from moral testimony, which are clearly accessible.³¹ Public reason liberals have raised fewer objections to relying on moral convictions than on religious convictions.³² Even so, moral testimony and religious testimony are epistemically symmetrical. To see this, bear in mind that moral reasoning often relies on testimony from others—from families, communities, teachers, parents, respected authorities and books. Our moral judgments seldom arise from pure reason; instead we form many moral beliefs based on the norms those around us already accept. Consider that most citizens cannot defend the moral claims they regularly rely upon, including those they rely upon in their political activities. We simply accept the moral judgments of others, or of authorities like parents, priests and friends.³³

Consider how our moral beliefs based on moral testimony acquire epistemic justification. One way to achieve justification is by believing the claims of those we take to be *reliable*. Reliability can be understood in several ways—that testifiers are perceptive, rational, knowledgeable, cool-headed, truth-tracking, etc. Generally we say that testifiers are reliable when they testify based on *good reasons*, that is, their testimony is epistemically justified. Testifiers testify based on good reasons when their testimony can be traced back to a justified judgment not dependent on testimony even if the judgment may be made by someone further “upstream” a testimonial chain. It seems clear that Sarah can justifiably affirm a moral proposition if she believes the moral testimony of John, a man she justifiably believes is reliable. She may believe he is reliable because his judgment is either not justified by testimony or can be traced back to such a belief in another person.

Let us accept that some moral beliefs derived from testimony are justified according to common evaluative standards. Presumably there are some widely shared criteria on which the reliability of a moral testifier can be judged. If John the Testifier is honest, well-informed, level-headed and (at least tacitly) employs a reasonable standard of evidence, Sarah probably justifiably believes that John is reliable; consequently, the moral beliefs Sarah forms from accepting John’s

³¹ Christopher Eberle makes such an argument, drawing an analogy between justifications derived from morality and those derived from religious experience. He claims that neither moral nor religious rationales are subject to independent confirmation and that there is no non-circular argument on behalf of the reliability of religious experiences or moral rationales. See Eberle 2002, p. 245.

³² As Thomas Nagel has claimed: “conflicts of religious faith fail this test [of common critical rationality], and more empirical and many moral disagreements do not.” Nagel 1987, p. 270.

³³ For more on trusting the testimony of our communities, see Reid 1983, pp. 281-2.

testimony will be justified. Others may well make similar judgments. Thus, while a moral reason like “The moral authorities in my life think X is morally wrong, so X must be wrong” may seem inaccessible, it appears justifiable via common evaluative standards. Therefore, such reasons are accessible when the relevant testifiers are honest, well-informed, level-headed, etc. We can tentatively conclude, then, that many moral judgments derived from testimony will be accessible at the right level of idealization.

A critic might argue that testimonial beliefs are redundant at the right level of idealization because agents will already have direct access to all the information they would learn from testimony. This is implausible. Evaluating the information communicated to an agent via testimony would be an epic undertaking. A plausible account of idealization does not employ ideal agents with god-like capacities, but ones that utilize cognitive processes similar to unidealized agents (Ch.9, II.i). Thus agents with limited cognitive faculties will be unable to process all the relevant information. Relying on testimony economizes on the costs of collecting information and processing it. A plausible idealization-value will therefore include testimonial beliefs, some of which may concern morality. Another problem needs addressing. Often we consider testimonial beliefs accessible because we can check the reliability of the testifier’s source. One might think checking reliability is difficult for moral testimony. However, the public reason tradition already supposes that moral reasoning is reliable in that it either tracks the moral truth or follows the correct procedures for good reasoning. Thus, if someone testifies about morality, checking the reliability of the testifier’s source requires checking the reasoning for flaws or for an appropriate degree of epistemic justification.

Reasons derived from religious testimony are accessible on analogy. Moral testimony is justifiably accepted if the relevant testifier is reliable. Suppose that the testifier has solid epistemic credentials because her testimony can be traced to a long and well-developed tradition of moral reasoning. Even this high standard of reliability will count many religious testifiers as reliable. For instance, the natural theological arguments discussed above can provide justified grounds on which to believe supernatural moral claims. Those who testify based on the reasoning of these natural theologians seem reliable as a result. To illustrate, the moral beliefs of many Catholics derive from their local priests. In seminary, these priests probably studied serious Catholic philosophers, including St. Augustine, St. Anselm, and St. Thomas Aquinas. As a result, these priests may have reasonable arguments for their positions, or know someone who does, even if those arguments are

flawed. In this case, the religious testimony of Catholics traces back to a reliable source, a source arguably more reliable than many sources of moral testimony.

Many will remain skeptical of the connection, but it is unclear where it fails. Moral testifiers are frequently embedded in communities and traditions of moral reasoning and are often reliable in the sense described. Arguably then, some reasonable people can justifiably trust their testimony. This trust will produce accessible testimonial beliefs about morality. Those who accept religious testimony are in a symmetrical epistemic position. Religious testifiers are often embedded within intelligent communities and rich traditions of theological and moral reasoning; further, at least some of them are reliable in the sense described. Those who believe religious testimony can develop justified beliefs based on that testimony. As a result, they will have accessible testimonial beliefs.

Consider an illustration. An economically underprivileged black man attends Dexter Avenue Baptist Church in Montgomery, Alabama in 1955; call him Claude. He hears that on December 1st, Rosa Parks refused to give up her seat on a Montgomery Bus thereby violating Alabama's Jim Crow laws. Claude's pastor, Dr. Martin Luther King, Jr., tells his church that they are going to boycott the Montgomery bus system until the buses are desegregated. He tells his congregation, full of men like Claude, that segregation is an affront to their humanity because God created men with equal dignity. Therefore, blacks are equal to whites. As a result, segregation laws degrade black people and must be resisted, even if it means going to jail. Claude accepts King's testimony, despite not entirely understanding King's reasoning. While he shares with King the sentiment that segregation is degrading, he trusts King's explanation as to why it is so. Claude next engages in a political activity—he tries to change the Jim Crow laws. He believes, based on the testimony of Martin Luther King, Jr. that the law should be changed and he therefore continually engages in political activity to force the city of Montgomery to desegregate its buses.

Suppose that Bobby approaches Claude and asks him why he is trying to change the law, and suppose that Claude responds, "Well, my pastor, Dr. Martin King, told me that the Bible says God created all men equal and that they shouldn't be treated unequally; so that's why I'm boycotting, because I believe him and I trust him." Imagine that Bobby is a traditional public reason liberal who then replies: "Claude, I'm sorry, but you have just given me an inaccessible reason. I'm not a Christian and I don't much trust pastors when it comes to morality. By offering such a reason, you're disrespecting me by asserting your authority to change the law without giving me a reason that I can access." I submit that not only is Claude not disrespecting Bobby, but that his reason is perfectly accessible to Bobby. Bobby's reaction to Claude seems obtuse and even bizarre. Bobby has

probably heard of Dr. King and if he is truly a public reason liberal, he will regard Dr. King's testimony as reliable from Claude's perspective. He will find that Claude has a justified belief and will therefore have to conclude that Claude's reasoning meets the accessibility requirement.

I have now reviewed several strategies that show religious testimony can satisfy the accessibility requirement. I conclude, therefore, that some reasons derived from religious testimony can satisfy the accessibility requirement. Public reason liberals cannot use the accessibility requirement to block reasons derived from religious testimony from entering into public justification. Since reasons derived from religious testimony are, again, the paradigmatic private reasons, the fact that accessibility cannot exclude them seems to deprive accessibility of its force, thus demonstrating that it fails to meet its aims.

IV: Modifying the Level of Idealization

If reasons derived from natural theology and religious testimony are accessible, then the accessibility requirement fails to perform its function of "[separating] the public wheat from the private chaff."³⁴ If accessibility fails to exclude what it is traditionally thought to exclude (even meant to exclude) it fails to achieve its aims. As such, accessibility should be jettisoned from conceptions of public reasons.

Two strategies can save accessibility: replacing accessibility with a related, but similar idea or changing the relevant level of idealization. The first option might involve replacing accessibility with, for example, external criticizeability. The second option might involve idealizing more or idealizing less. Regardless of strategy, however, the alternative must block reasons public reason liberals regard as suspect without blocking them merely for that reason. The public reason liberal must exclude reasons she regards as problematic from the point of view of public reason while treating the reasoning of each person impartially.

The search for a new understanding of accessibility seems like a dead end. While standards abound in the literature, their similarity to accessibility renders them vulnerable. The standards all appear to require that public reasons are justified according to common evaluative standards at the right level of idealization. We have already reviewed three distinct conceptions of accessibility in the literature (Ch.5, II.i) and derived the accessibility requirement by generalizing from their common

³⁴ Eberle 2002, p. 14.

elements.³⁵ For this reason, it is hard to imagine a requirement sufficiently similar to accessibility to do the job it cannot.

The defender of accessibility must therefore develop a new conception of idealization that is either more or less demanding. From the preceding argument, it appears that conceptions of idealization closer to full rationality and full information will change little. The religious reasons in question do not require significant rational capacity or enormous amounts of information to access, evaluate or independently confirm. They only require a basic familiarity with the arguments that support them. As a result, it is not clear how additional information and rational capacity could render religious arguments inaccessible. If Section III's argument succeeds, any plausible conception of idealization will contain testimonial truths, and if Section IV's argument succeeds, even modest rational capacities can render natural theological arguments accessible. A critic might argue that more information and rational capacity will render some arguments inaccessible. For instance, philosophers sometimes refuse to accord claims justification that non-philosophers regard as justified. In some cases, philosophers reach a degree of understanding that shows the claim to be unjustified. Consequently, a higher level of idealization might make some claims inaccessible. Yet while higher levels of idealization may render some claims inaccessible, we have little reason to think that they will render arguments from natural theology and religious testimony inaccessible in general. As we saw above, too many thoughtful people embrace these arguments to reasonably maintain that at high levels of idealization, none of these arguments will be regarded as justified. Here's why: if idealized leads to the renunciation of arguments from natural theology and religious testimony, then entire worldviews and religions become irrational to affirm. Construing idealization as showing that entire religious traditions with thousands of years of theological discourse are unreasonable seems to deny reasonable pluralism. If public reason liberals decide to idealize in this way, then they undercut one of their foundational commitments. It may well be that on a fully informed and fully rational conception of idealization, a vast swath of religious beliefs (if not all religious beliefs) will be rejected. But if this is so, then such a degree of idealization is off the table for public reason liberals.

In response to the failure of increasing idealization, public reason liberals might opt for a less demanding conception of idealization than those ordinarily employed. The thought might be that by reducing idealization levels, we might find that religious reasons are inaccessible while secular

³⁵ I take it for granted that the individuating features of the distinct accounts of stability will not alter the assessment of natural theology or testimony mentioned above.

reasons are accessible. One such idealization level is known as “populist.”³⁶ Fully populist conceptions of idealization require no idealization at all, but they face significant problems. If public reason liberals adopt populist idealization-values, public justification will be captive to bad information and poor reasoning.³⁷ However, public reasoning subject to bad information and poor reasoning is arguably not public reasoning at all. Recall that we have understood public justification in terms of open justification (Ch.4: II.i), which ascribes reasons to agents that can be derived via sound inferences from their current belief-value set. A populist conception of idealization is incompatible with a commitment to open justification because open justification permits reasons to be ascribed to individuals based on inferential connections that they may not ordinarily make or even be able to make. Thus, populist conceptions of idealization appear ineligible as part of a conception of public reasons. Populism therefore cannot help accessibility achieve its aims qua being a conception of public reasons since populism is ineligible on all public reason views.

Neither high nor populist conceptions of idealization help accessibility achieve its aims. The only option left is to adopt a “near-populist” idealization value, one below the commonsense level appealed to in Sections III and IV and above populism. By adopting a “near-populist” idealization value, public reason liberals idealize only mildly. Depending on the details, a defender of accessibility could argue that many mildly idealized citizens could not see natural theological arguments as justified according to common evaluative standards. While perhaps true, this point proves too much. Consider two examples. First, the particulars of climatology required to form cogent global warming policy are far more complicated than the information required to evaluate arguments rooted in religious testimony. Public reason liberals will want enough idealization to employ climatology (or reliable testimony about climatology); if so, they will have trouble excluding arguments rooted in religious testimony and natural theology. Substantially less cognitive capacity is required to assess religious arguments than arguments based on climate science. Even if the reliability of a testifier can be established within climate science and thereby supply the relevant information, it is, again, unclear how religious testifiers can be excluded. Second, many public reason liberals are ardent opponents of teaching intelligent design in the classroom alongside evolution. To justify excluding intelligent design, public reason liberals could show that at the right level of idealization parents and teachers (or the relevant authorities) will recognize that intelligent design is bad science. But the ability to

³⁶ Gaus 1996, pp. 130-136.

³⁷ *Ibid.*, pp. 130-131.

identify good science goes beyond the ability to assess arguments based on natural theology and religious testimony.

Consequently, near-populist interpretations of accessibility seem to be dead ends since they are implausibly restrictive. First, they restrict first-personal liberty because they prevent citizens from offering basic scientific reasoning in public, something reasonable and well-informed citizens will wish to do. Further, many will be committed to offering such reasoning based on deeply affirmed principles, such as placing a high value on scientific education. Second, by restricting the ability to discuss vital information relevant to the goodness of a political proposal, conceptions of accessibility with near-populist idealization values seem to be poor expressions of a commitment to public justification. Somewhat like populist idealization values, they hold public reasoning captive to unacceptably poor information and reasoning.

Sections II, III and IV show the accessibility requirement either fails to achieve its aims or is excessively restrictive in first and second-personal terms. We can conclude that conceptions of public reasons containing the accessibility requirement rank below conceptions without it. If two conceptions are equivalent in every respect save accessibility, the conception without accessibility is superior to the conception that contains it. Thus, while some conceptions of public reasons with accessibility may be superior to some conceptions without it, for each conception of public reasons with accessibility there is likely a superior conception otherwise similar that lacks it. Conceptions of accessibility with relatively more idealization cannot achieve their aims. But conceptions of accessibility with relatively less idealization fail both the first-personal and second-personal desiderata.

V: Shareability

Recall that shareability adds a shared reasons requirement to accessibility's shared evaluative standards requirement. Thus, shareability requires that both evaluative standards and reasons be shared. Shareability will therefore be more restrictive than accessibility when the idealization value is held constant. However, the reasons that count as accessible seem to vary quite a bit depending on the idealization value adopted. This does not seem to be the case with shareability. At low idealization values, citizens clearly share few reasons, so shareability is restrictive; but we shall also see that it is restrictive at high idealization values. Thus, shareability is always restrictive. It always sets a large set of reasons aside. Since shareability is restrictive, conceptions of public reasons that

contain shareability seem inferior to conceptions that lack it based on both the first and second-personal desiderata.

The restrictiveness of shareability is not hard to demonstrate. At populist or near-populist idealization values, shareability will exclude a vast number of reasons. Due to reasonable pluralism, citizens' reasons differ and do so justifiably. If intelligibility is the appropriate epistemic standard for public reason liberalism, individuals will justifiably affirm distinct reasons as they arrive at those reasons from distinct beliefs and values. A mildly idealized shareability requirement bases politics (largely) on reasons that citizens currently share. Surprisingly, a handful of public reason liberals do not seem to mind such a stringent requirement.³⁸ Nonetheless, such interpretations of shareability are quite restrictive. These shareability requirements seem to impose just the sorts of first-personal restrictions that raise the ire of integralists. As principles of exclusion, citizens cannot appeal to unshared reasons as defeaters. This appears to render inappropriate many arguments traditionally recognized as sound by the liberal tradition, such as arguments for religious accommodation.³⁹ As a principle of restraint, these shareability requirements seem quite restrictive of integrity (Ch.2, II.iii-II.vi). Shareability requirements are also restrictive on the second-personal metric because they substantially reduce the scope of reasonable pluralism with respect to the sorts of diverse reasons that bear on exclusion and restraint.

The same problems hold at high levels of idealization. To see this, suppose that we radically idealize members of the public by giving them near perfect information and cognitive abilities; in this case, they will either continue to disagree about reasons or not. If idealized citizens affirm only the same reasons, then adopting such a conception effectively rejects reasonable pluralism.⁴⁰ To assume that all citizens' reasons are shared is *just to deny* reasonable pluralism. But what if highly idealized citizens have different reasons? Many in the public reason tradition will find this hard to accept, but the burdens of judgment will still apply to highly idealized citizens.⁴¹ Recall that Rawls recognized six burdens of judgment as sources of reasonable disagreement: (a) evidence is conflicting and complex and so hard to assess, (b) people disagree about the weight of their shared considerations, (c) political concepts are subject to hard cases and indeterminacy and so require

³⁸ See Bohman and Richardson 2010 and Schwartzman 2011 for recent defenses of shared reasons standards.

³⁹ A consensus liberal might reply that there is a shared "liberty of conscience" reason. I do not think this response is adequate since it fails to base liberty of conscience on the unshared reasons of the public, which is likely where the primary rationale of liberty of conscience lies. However, I cannot pursue this matter in the dissertation at length as it requires a sociological examination of different historical objections.

⁴⁰ See Rawls 1971, pp. 15-19. In Ch.7, we will examine how and why Rawls's theory of public justification lacks a shareability requirement despite his conception of idealization.

⁴¹ See Rawls 2005, pp. 54-58 for a discussion of the burdens of judgment and their relationship to reasonable pluralism.

individual judgment, (d) our life experiences shape how we assess evidence and weigh values yet our experiences always differ, (e) different kinds of normative considerations lie on either side of an issue, making a general assessment difficult and (f) due to the opportunity costs of implementing values, we must make hard choices without clear direction.⁴² None of these conditions disappear at high levels of idealization. We can still disagree about the relevant evidence and weights of considerations. Our concepts will still face indeterminacy and our life experiences will still shape our perspectives. Different normative considerations will still be hard to reconcile and we will still have to choose which values to implement. If so, highly idealized individuals will disagree. In fact, since they have more information and greater cognitive capacities, members of the public may disagree *more* than their counterparts. If so, shareability prohibits appeal to a huge number of reasons. Accordingly, shareability still seems restrictive. Thus shareability seems restrictive whether it decreases or increases disagreement among members of the public.

A proponent of shareability can counter by claiming that despite reasonable disagreement, the set of shared reasons is still large. Thus, while John and Reba may disagree about whether considerations A through M are reasons, they may still share reasons N through Z. Consequently, reasonable disagreement could remain significant, despite an expansive realm of shared reasons. Nonetheless, shareability should still be rejected since it is not entailed by the Public Justification Principle. It still appears that conceptions of public reasons that lack shareability are superior to those that accept it due to shareability's restrictive nature. If so, the fact that the set of shared reasons is large cannot save shareability. The restrictions imposed by the shareability requirement still seem to provide decisive reason to reject it. Since conceptions of public reasons that lack shareability fail to imply principles of restraint, there appear to be many conceptions of public reasons that are superior to shareability on first and second-personal grounds.

To complete the case against shareability, the three arguments for shareability discussed in Ch.5, III.ii must be assessed. The first argument claims that shareability is entailed by a commitment to respect for persons: to treat others with respect, we may only offer them shared reasons. But these arguments from respect for persons to shareability are identical to those offered on behalf of accessibility. If public reason liberals take both shareability and accessibility to follow from respect for persons, then they are wrong about *both claims* since the same arguments are used on behalf of two distinct requirements. A critic could reply that the arguments from respect for persons entail accessibility by showing that evaluative standards must be shared and entails shareability by showing

⁴² Ibid., pp. 56-57.

reasons must also be shared. The inconsistency dissolves because the arguments vindicate accessibility by vindicating shareability. Even so, mainstream public reason liberals have been unclear about whether respect for persons requires shared reasons, despite agreeing that it requires shared evaluative standards. The ambiguity still suggests that the conceptual argument fails.

The second argument for shareability holds that all public reasons are *by definition* shareable reasons. Scanlon, Korsgaard and Nagel hold that a consideration counts as a reason for X only if it counts as a reason for any person placed in identical circumstances. This claim is either false or trivial depending on what “in identical circumstances” means. First, given intelligibility, any consideration that can be recognizably inferred from one’s current belief-value set is a reason. If “in identical circumstances” means “if two people have identical beliefs and values,” then the claim is trivial; of course two people with identical beliefs and values will have identical reasons. Shareability interpreted in this way is redundant. In contrast, if “in identical circumstances” is more restrictive than having identical beliefs and values, Scanlon, Korsgaard and Nagel are wrong about what counts as a (public) reason. If a reason is bona fide whenever it is intelligible, unshared reasons are *bona fide*.

The third argument holds that publicity requires shareability. A society possesses publicity when it is aware that its political institutions are justified and aware of the reasons why. Shareability is thought to ease the evaluation of rationales (Ch.3, III.iii). If citizens share considerations and evaluative standards then publicity is easier to recognize. The publicity argument claims no direct entailment from reasoning publicly to shareability; it must therefore be evaluated according to whether its benefits are sufficient to outweigh the disadvantages of conceptions of public reasons that include shareability. The benefits are probably insufficient for reasons I have already suggested (Ch.3, III.iii). Permitting unshared reasons into public dialogue and allowing them to play a justificatory role will not prevent a society from achieving publicity since the milder form of “rule” publicity I defended can perform the same role as the “full” publicity required by mainstream public reason liberals. We can conclude that public reason liberals should reject shareability as well as accessibility.

VI: Symmetry

Symmetry requires that the same restrictions apply to reasons offered for coercive proposals and to those offered against coercive proposals. For instance, if shareability holds for reasons to reject proposals, then symmetry mandates that shareability apply to reasons that support those proposals.

One might wonder whether symmetry has any benefits. There is little to say, since no such benefits have been advanced. Symmetry may have advantages, but it is unclear what they would be. Instead, symmetry seems to only complicate matters depending on which requirements it is applied to. To see this, first notice that intelligibility is automatically symmetrical as it is required by the very idea of reasoning publicly. Since intelligibility is symmetrical without the symmetry requirement, *symmetry has nothing to do*. But notice that if the arguments above succeed, we have decisive reason to reject the accessibility and shareability requirements. Consequently, since accessibility and shareability should be rejected, symmetry cannot be applied to them. Accordingly, either symmetry is superfluous, as it is for intelligibility, or else there is no standard to which it can be applied. Since the symmetry requirement is presumed to do *something* it fails to achieve its aims. We can conclude, therefore, that symmetry should be rejected.

VII: Sincerity

I think that the sincerity requirement can be understood in terms of the above four requirements. Specifically, the sincerity requirement can be understood as the application of a set of requirements to a principle of restraint, that is, to what reasons may be advanced and acted upon in the public square. If accessibility is the only relevant requirement, then a sincerity principle requires only offering accessible reasons in public political argument. Here are two sincerity requirements, one from Gaus and the other from Jonathan Quong:

Gaus: Betty's argument justifying N to Alf is sincere if and only if (1) she is justified in accepting N; (2) she has a justified belief that N is justifiable in Alf's system of reasons and beliefs.⁴³

Quong: A may only endorse X if the following are true (and vice versa for B): (1) A reasonably believes he is justified in endorsing X, (2) A reasonably believes that B is justified in endorsing X and thus (3) A may only (in the political domain) offer arguments in favor of X to B that he reasonably believes B would be justified in accepting.⁴⁴

⁴³ Gaus 1996, p. 140.

⁴⁴ Quong 2010, p. 266.

Gaus's principle of sincerity holds that B can only offer A reason R for proposal P if P is justified to B and B is justified in believing that P is justified to A. Quong's principle is nearly identical, despite using the distinct property of reasonableness. Thus both principles require that arguments in the public sphere be offered in an honest attempt at public justification. If B does not believe P is justified to A then by offering R, B appears to deceive or manipulate A. Conversely, if B does not believe that P is justified to B, then by offering P to A she also seems deceptive or manipulative. B is attempting to convince A to endorse P despite believing that it is unjustified.⁴⁵

To decompose sincerity into other requirements, let's ask why anyone should care about being sincere. The common answer is that sincerity is required by respect for persons. If we believe that coercion must be publicly justified out of respect for others but we ask others to accept coercion for bad reasons, we disrespect them by our own lights. The problem with the common answer is that the Public Justification Principle by itself is ambiguous between different conceptions of public reasons. It is thereby ambiguous between different conceptions of the evaluative standards required to count something as a reason. For instance, accessibility and shareability require evaluative standards to be shared, whereas intelligibility does not. If accessibility holds, then *what it means* to be sincere will be to offer arguments for proposals that one believes are justified according to common evaluative standards. Similarly, if only intelligibility holds, then *what it means* to be sincere is to offer arguments for proposals that one believes are justified according to the evaluative standards of one's discussant and one's own evaluative standards. Sincerity admits of different versions depending on which requirements we already endorse. Quong's principle seems to suppose common evaluative standards.⁴⁶ If so, he has merely applied (at least) an accessibility requirement to restraint.⁴⁷

Does the best conception of public reasons contain a sincerity requirement? It initially appears to, since the Public Justification Principle contains the intelligibility requirement. Whether a sincerity requirement is justified will depend on (a) whether intelligibility entails a principle of exclusion and (b) whether the principle of exclusion entails a principle of restraint. One could object that whether a conception of exclusion or restraint is sound cannot depend on entailment. After all,

⁴⁵ Gaus has maintained in exchange that his sincerity principle concerns what reasons are justificatory, not which reasons can be offered in speech acts. If so, then his sincerity principle is not our target. Instead, I take a justificatory principle of sincerity to have already been captured by my definition of first and second-personal reasons in Chapter 4.

⁴⁶ While Gaus's principle may appear to imply common evaluative standards, Gaus in fact embraces a form of cognitive relativism. See Gaus 1996, pp. 38-44.

⁴⁷ Quong clearly intends for sincerity to apply to restraint, given his claim that "A may only (in the political domain) *offer arguments* in favor of X to be that he reasonably believe B would be justified in accepting." Quong 2010, p. 266. Emphasis mine.

the entire strategy of the dissertation is to select a conception of public justification from an eligible set of conceptions using several desiderata, none of which are entailed by the Public Justification Principle. Nonetheless, intelligibility entails a principle of exclusion by holding that a reason can only justify coercion if it is intelligible to members of the public. Arguably any unintelligible reason should be excluded from the process of public justification. An unintelligible reason cannot serve as a basis of coercion at all, so it should not even enter the process of public justification as it does not count as a public reason. Further, institutions should be designed (in as much as they can) so that unintelligible reasons do not significantly move a society towards publicly unjustified laws. The rule of exclusion is therefore this: no unintelligible reason can serve as a legitimate basis of coercion.

We must now ask whether the foregoing principle of exclusion entails a principle of restraint. As we have seen (Ch.2: II.ii), a case must be made to move from exclusion to restraint. If one goal of public justification is to prevent nonpublic reasons from being bases of coercion, then the best principle of restraint is one that, in practice, tends to bring this about. Selecting such a principle seems to require complex sociological judgments, though. For instance, Gaus has maintained that principles of restraint that bar wrongful advocacy often oversimplify politics in important ways. To illustrate, Gaus asks us to consider the following case:

Alf's Immoderate Proposal: Alf believes that, ideally, a center-left free market oriented government would be publicly justified. However he is convinced that free trade is the most important issue today—though he would certainly endorse government programs to give significant aid to displaced workers. But Alf reasonably thinks that, in the current shrill political climate, nuanced proposals get lost in the noise; if he advocates government programs to aid displaced workers, his speech will be coded as “anti-trade.” So Alf publicly advocates a radical free market approach, always talking about the benefits of free trade and never its shortcomings, though he sincerely hopes that such an approach is not instituted, nor, given the need for political compromise, does he expect it to be. Alf thinks that the most probable result of this advocacy will be a final policy a wee bit closer to what is publicly justified.⁴⁸

Gaus avers that Alf does not act objectionably given his political circumstances; I agree. What is politically permissible must be sensitive to social and political context. Thus whether reason X

⁴⁸ Gaus 2010, p. 27

becomes a basis for coercion depends on a host of sociological factors which a principle of restraint could affect for the better or for the worse.

One might think that a principle of restraint concerns ordinary political etiquette, i.e., how to treat people with respect in ordinary interactions by not being deceitful, rude and domineering. While these norms are important, they are not part of public reason *as such*. There are many reasons to endorse such norms that do not derive from the Public Justification Principle. It is likely that such commonsense standards of civility have been mixed with public reason because both are implications of treating others respectfully. If I respect you, I will publicly justify coercing you; similarly, if I respect you, I will not try to deceive you. But being sincere is not required by a commitment to public justification. We can, after all, imagine a large number of citizens making immoderate and mildly deceptive proposals in perfectly permissible ways (as understood by the public justification principle).⁴⁹ We must not confuse what is required by public reason with what is required by being a decent human being in a general sense.

A distinction should move the discussion along. Restraint can apply to two objects: *reasons* and *proposals*. Public reason liberals may advocate what I shall hereafter call “reason-restraint,” restraints on what reasons may be offered, or they may advocate what I shall refer to as “proposal-restraint,” restraints on which proposals may be advanced. The foregoing shows that there is little reason to embrace reason-restraint. But what about proposal-restraint? In lieu of further discussion, a limited principle of proposal-restraint likely falls out of the Public Justification Principle. Citizens who have a substantive causal impact on whether others are coerced should not advance proposals that they believe are publicly unjustified. However, this principle does not follow from the principle of exclusion discussed above. Rather it derives directly from a commitment to public justification. Citizens of influence should act so as to prevent unjustified proposals from becoming law because they should try to prevent wrongs when they can and coercion on the basis of an unjustified proposal is wrong.⁵⁰

Since we have little reason to embrace reason-restraint, we have little reason to embrace a sincerity requirement since a principle of proposal-restraint cannot be a principle of sincerity. Gaus has emphasized not getting too caught up on the term “sincerity”.⁵¹ But this only illustrates that

⁴⁹ If we consider the public justification of morality as a whole, then these commonsense norms will come under public reason, but for now we focus on politics.

⁵⁰ I will not discuss proposal-restraint further in the dissertation.

⁵¹ Gaus 1996, p. 140.

sincerity is a red-herring. It dissolves when we look at the foundational requirements on public reasons demanded by the Public Justification Principle.

VIII: Conclusion

This chapter has argued that the accessibility, shareability and symmetry requirements should be rejected. Since the Public Justification Principle is ambiguous between the standard conception of public reasons and less restrictive alternatives, we should embrace the less restrictive alternatives and reject the more restrictive standard conception. Thus, by demonstrating that accessibility (on some interpretations) and shareability have first and second-personal costs, we can show that conceptions of public reasons containing both requirements are inferior to those that lack them. By demonstrating that accessibility (on other interpretations) and symmetry fail to meet their aims, we can show the same.

The strategy of Part I pays big dividends. Now that we have explained the first and second-person duality within the foundational commitments of public reason, we have a compelling method of selecting among conceptions of public reasons. The strategy of Part I recommends economizing on first and second-personal costs when selecting a conception of public reasons. To defeat the standard conception, we need merely point to its first and second-personal costs and dispel the appearance of benefits; that is what Chapter 6 has done.

Making clear distinctions among the various requirements has an added advantage: it allows us to sort through a number of common problems in the public reason tradition and show how they relate to one another. For instance, we saw in Section VII that sincerity requirements are in fact derived from a conjunction of a principle of restraint and some other requirement(s). We can determine whether public reason requires a particular form of sincerity by determining whether restraint is required by the public justification principle and what sort of evaluative standards are appropriate to a conception of public reasons. When combined, they yield a principle that determines how to be sincere.

This chapter only argues against the standard conception of public reasons by showing that conceptions of public reasons lacking accessibility, shareability and symmetry requirements rank above those that contain them. In Chapter 7, I will develop and defend one of these conceptions.

Chapter 7: Convergence

This chapter builds on Chapter 6 by defending convergence. Recall that convergence embraces the intelligibility requirement but rejects the accessibility, shareability and symmetry requirements embraced by the consensus conception. Chapter 6 argued that conceptions of public reasons lacking accessibility, shareability and symmetry requirements are superior to conceptions that contain them. Consequently, convergence is the *prima facie* preferred conception.

What else must be said? We must still show that convergence performs adequately on the first and second-personal criteria developed in Part I. If convergence imposes too many first and second-personal costs, integralist liberals will still have good reason to reject it. If convergence and consensus jointly exhaust eligible interpretations of reasons, then integralists will have sound reason to reject public reason liberalism. The first task of the chapter, therefore, is to show that convergence imposes modest first and second-personal costs. The second task of the chapter is to address three other objections, one from a Rawlsian perspective and the other two from recent works. The third task of this chapter is to explain how convergence solves the central problem of the dissertation. Section I takes up the first task, Sections II-IV defend convergence against objections and Section V completes the final task of the chapter but ends by showing that convergence alone cannot resolve the role of religion debate.

I: Convergence from the First and Second-Person Points of View

Convergence is defensible only if it imposes modest first and second-personal costs on citizens. Recall that the first and second-personal standards are understood in terms of integrity costs and restrictions on reasonable pluralism. The metrics only provide a method of selecting among eligible interpretations of the Public Justification Principle. But integralist liberals might want the integrity of citizens to be preserved to a greater degree. Showing that convergence is superior to consensus does not thereby show that convergence is sufficiently friendly to integrity to give citizens reason to sign on to the demands of public reason. What we must show, following Chapter 4, is that the first and second-personal costs of convergence are sufficiently low to show that citizens of faith have reason to impose these costs on themselves, thereby reconciling the moral demands of public reason liberalism with integrity and leaving citizens of faith as free as before. One way to do this is to show that convergence imposes fewer costs upon integrity than even integralists' own standards of

restraint. Integralist liberals will therefore have no complaint. To do so, let us contrast convergence with a principle of restraint defended by one of public reason liberalism's most perceptive integralists, Christopher Eberle. Eberle rejects the principles of restraint associated with public reason liberalism (Ch.2, II.vi) but he advocates a kind of restraint in its place, "The Ideal of Conscientious Engagement".¹ If convergence imposes fewer integrity costs than the Ideal of Conscientious Engagement, we will have a strong case that convergence imposes an acceptable level of integrity cost even for integralist liberals. If so, integralists will have no objection to my core claim that a convergence conception of public reason is one according to which citizens of faith can *self-legislate* and thereby retain their freedom within the bounds of public reason.

Someone who pursues Eberle's ideal acts as follows:

- (1) She will pursue a high degree of rational justification for the claim that a favored coercive policy is morally appropriate.
- (2) She will withhold support from a given coercive policy if she can't acquire a sufficiently high degree of rational justification for the claim that that policy is morally appropriate.
- (3) She will attempt to communicate to her compatriots her reasons for coercing them.
- (4) She will pursue public justification for her favored coercive policies.
- (5) She will listen to her compatriots' evaluation of her reasons for her favored coercive policies with the intention of learning from them about the moral (im)propriety of those policies.
- (6) She will not support any policy on the basis of a rationale that denies the dignity of her compatriots.²

Living up to Eberle's ideal might be quite onerous. First, a citizen who fails to live up to the ideal can be properly criticized for failing to do so, since Eberle claims that a citizen "who fails to satisfy that ideal is, *ceteris paribus*, the object of reasonable moral criticism."³ As we have seen (Ch.4, I), moral criticism is a form of interference and imposes costs on the liberty of citizens. In this way, citizens face alienation from their community if they do not conscientiously engage other citizens. Recall Paul Weithman's similar complaint against the Rawlsian duty of civility. Weithman worries that the duty of civility requires citizens to develop significant critical capacities in order to permissibly engage in public political action (Ch.2, II.v). Eberle's ideal is vulnerable to the same

¹ Eberle 2002, p. 104.

² Ibid.

³ Ibid., p. 105.

concern. To discharge their public duties, citizens must reason at a high degree of sophistication and effectively gather evidence. Further, they must have developed communication skills and sufficient moral virtue to refrain from supporting policies that violate the dignity of their compatriots. If citizens do not expend the costs required to acquire these abilities and act on them, they can be ostracized.⁴

To vindicate convergence, we need merely show that it imposes fewer restraints than the ideal of conscientious engagement. Given the restrictions the ideal of conscientious engagement imposes of citizens, this is an easy task. As we saw in Chapter 2 (Ch.2, II.i), the Public Justification Principle only requires restraint by means of consensus. Convergence is far more permissive with respect to the reasons it regards as justificatory. It recognizes all *intelligible* reasons as justificatory. Consequently, it cannot ground the substantial restraint on reasons defended by consensus liberals. That said, I have indicated (Ch.6, VII) that citizens must sometimes engage in *proposal-restraint*, restraint with respect to the advocacy of coercion. Citizens should not advocate coercion they believe cannot be justified to others if their advocacy can significantly impact the outcome. I did not flesh out the proposal in detail, but I do think convergence entails some principle of this sort. Even so, proposal-restraint imposes far fewer integrity costs than Eberle's Ideal of Conscientious Engagement. Proposal-restraint imposes no restrictions on the reasons citizens may act upon and offer in public political life. Instead, proposal-restraint merely requires that citizens of influence refrain from supporting laws that they believe are unjustified. Integrity costs may yet arise as one may feel conscience-bound to support a law regardless of whether it can be publicly justified, but again, these integrity costs are far less significant than those imposed by Eberle's ideal. Eberle presumably believes that citizens remain free even when obeying the Ideal of Conscientious Engagement. If so, it is hard to see why they would be less free if they embraced the convergence conception of public justification. It is important to drive home this point: *integralist concerns simply do not apply to convergence liberalism*. All the complaints discussed in Chapter 2 address the mainstream conception of public reason liberalism, not the convergence view. Consequently, a convergence conception of reasons promises to avoid integralist criticisms and answer the powerful integrity objection to public reason liberalism. If I am right that convergence imposes more modest costs to integrity than Eberle's ideal, the first-personal metric shows that convergence is superior to Eberle's

⁴ Acquiring these abilities is typically not required by most real-world individual ideals, so the burdens will almost always be seen as at least an inconvenience, but will often be seen as burdensome interruptions of one's plan of life.

standard and the consensus conception of public reasons. While these considerations do not vindicate convergence, they create a strong presumption on its behalf.

II: Rawls's Synthesis

Since convergence is presumptively superior to consensus as a conception of justificatory reasons, we should now consider what the consensus liberal will say in response. Perhaps our first step should be to address John Rawls's view, which is a kind of hybrid theory of political justification that combines consensus and convergence. To illustrate, recall the three-part structure of political justification in Rawls's *Political Liberalism*.⁵ (1) The first stage, the original position, generates a freestanding political conception of justice, a module that can fit within the comprehensive doctrines of all reasonable persons. In this stage of justification, the parties to the original position reason in the same fashion due to the stringent restrictions on information and reasoning behind the veil of ignorance, for "the veil of ignorance makes possible a unanimous choice of a particular conception of justice."⁶ The parties' reasoning does not draw on their comprehensive doctrines. This "pro tanto" stage of justification might be understood as an extreme consensus stage of justification, where parties to the original position reason on precisely the same basis. (2) In the next stage, full justification, citizens of the well-ordered society test the political conception to see if it can fit within each reasonable comprehensive doctrine. In this stage, the idea of a modular conception of justice becomes important. One reason for this is that the political conception "gives no guidance" about how it is to be embedded within citizens' comprehensive doctrines.⁷ In the full justification stage "it is left to each citizen ... to say how the claims of political justice are to be ordered, or weighed, against nonpolitical values. The political conception gives no guidance in such questions"⁸ Citizens of a well-ordered society must complete full justification themselves. We might understand full justification as the convergence stage of justification.

While Rawls thinks that the first two stages of political justification are necessary to publicly justify a political conception of justice, the convergence theorist disagrees, maintaining that full justification—convergence—is the only normatively significant stage of political justification. If

⁵ For the most developed articulation of Rawls's three stages of justification, see Rawls 1995, Rawls 2005, pp. 385-394. For Habermas's critical remarks see Habermas 1995, pp. 119-122, 109-131, esp. 119-122.

⁶ Rawls 1971, p. 121.

⁷ Rawls, *Ibid.*, pp. 386-387.

⁸ *Ibid.*

political institutions are justified to each person on the basis of her own reasons, political justification has been achieved. The convergence theorist wonders why political theory cannot do away with the first stage. While *pro tanto* justification might be employed as a useful heuristic for locating a set of potentially justified principles, it would function only as a discovery procedure to uncover points of convergence. There is ample reason to suspect that such a discovery procedure could successfully perform its theoretical function. Convergence reasons are a superset of consensus reasons. The convergence theorist can draw on *all* of the resources available to the consensus theorist.

Rawls might reply that the consensus or the free-standing justification is required to avoid indeterminacy in determining a conception of justice. As Rawls notes,

The veil of ignorance makes possible a unanimous choice of a particular conception of justice. Without these limitations on knowledge the bargaining problem of the original position would be *hopelessly* complicated. Even if theoretically a solution were to exist, we would not, at present anyway, be able to determine it.⁹

This is to say that members of the public cannot settle on clear principles of justice unless they begin political justification with a focus on developing a free-standing conception of justice. Rawls wants the original position to locate a determinate set of principles.¹⁰ Without determinacy, the original position seems to fail to identify the principles of justice as it might generate an unmanageably large number of conceptions of justice. Citizens' reasons are so varied that locating and selecting among the potential convergence points might be impossible. Rawls responds to this challenge by building substantive constraints into the original position to reduce the set of potentially justified principles to a manageable level. This is one *point* of the veil of ignorance.

Rawls's counter to mere convergence is that substantive constraints on permissible reasons are needed to manage indeterminacy and produce agreement on specific, concrete principles of justice. Yet why should anyone think that the set of potential political principles would be very large? A Rawlsian can argue in the following way: all members of the public have an interest in securing the gains of social cooperation. But there are many different social rules that can allow people to achieve some of these benefits. Rawlsians will worry that a mere convergence conception of public

⁹ Rawls 1971, p. 121. Emphasis mine.

¹⁰ In his early work, Rawls wants the parties to the original position to hear complaints until everyone is "roughly of one mind" as to how complaints are to be judged. See Rawls 1957, p.656.

justification will not be able to differentiate among these many principles. This worry will only be amplified by the fact of reasonable pluralism, which suggests that disagreement about principles of justice is a fundamental feature of any free society. Disagreement about the correct principles may be enormous and although all have an interest in *some* solution to the problem of social cooperation, they may not be able to agree on which one is best.

The convergence theorist can respond to Rawls's challenge in two ways: (a) argue that the set of potentially justified principles is not unmanageably large and (b) argue that parties selecting a conception of justice can publicly justify a decision procedure by which political principles may be selected. Option (a) may seem hopeless. As we have seen, reasonable pluralism may reveal many potentially justified principles. To pursue option (a), the convergence theorist must introduce new conceptual resources to reduce the number of potentially justified political principles. She can begin by reconsidering the place of the symmetry requirement within public reason liberalism. Again, the symmetry requirement holds that the restrictions that apply to reasons to propose coercion must apply to the reasons to reject coercion and vice versa. Suppose that the convergence theorist rejects symmetry as I have suggested. In this case, all public reasons could enter into public justification as potential defeaters for coercion. In other words, these diverse, comprehensive reasons could show that many potential political principles cannot be conclusively justified. Convergence reasons will frequently be able to undermine more proposals than consensus reasons since the set of convergence reasons is larger. Consequently, the convergence liberal could reduce the set of potentially publicly justified principles simply by relaxing the symmetry requirement. Individuals' private reasons are extraordinarily diverse; many different opportunities to pass coercive proposals could be shown to be unjustified. This suggests that a real danger of (asymmetric) convergence is that it will fail to generate *any* convergence points. Christopher Eberle has raised this concern, which we examine in Section III.

Option (b) involves two steps. First, it gathers the potential convergence points into a set and attempts to rank them according to what citizens have most reason to endorse. The list is comprised of all political principles citizens regard as an improvement over no principle. The second step requires choosing a decision procedure that can select a member of the set. Once a decision procedure is publicly justified, a society can select political principles. Let us unpack this a bit with an example. Suppose Alpha Centauri is trying to decide which principle of distributive justice to adopt. All three leaders of its citizens' factions agree that some sort of care should be given to the least advantaged, but they disagree about how much. Senator Arthur believes that distributive justice

requires providing sufficient resources so that the least advantaged never have to work to make their way through life. Senator Benedict believes that sufficient resources should be provided to the least advantaged until they reach adulthood and then distribution should be sensitive to their attempts to find employment. Senator Cyril believes that sufficient resources should only be guaranteed to those who fall into poverty through no fault of their own. However, Alpha Centauri at present provides no safety net of any kind. All three faction leaders think that any of their proposals is superior to their present circumstances. Any of these principles can be publicly justified to all three factions in comparison with their present circumstances. Thus in the first step, the three factions list their preferred principles, the Arthur principle, the Benedict principle and the Cyril principle, in an order that satisfies them all. They realize after some debate that they cannot convince the other two factions that their principle is best. So they select a majority voting rule to decide among the three principles in the list. They all regard the majority voting rule as fair. In this way, all three factions on Alpha Centauri can use convergence reasoning, face the prospect of a large set of potentially justified principles, and use a justified voting rule to arbitrate among them.

This strategy is essentially the solution to the problem of indeterminacy proposed by the social contract tradition as a whole; particular structures of government are justified on the grounds that they help to resolve disputes about rankings. Hobbes pursues this strategy by defending the use of a sovereign power and it is on this basis that Locke and Rousseau recommend democracy.¹¹ But the point bears expanding. Rawls's solution to the problem of selecting from a multiplicity of potentially justified principles of justice is primarily theoretical in the following sense. The constraints by which parties to the original position are led to select a single set of principles are built into the rationality of the parties. In other words, Rawls's parties to the original position are not *bargainers*. They all think in precisely the same way. As such, there is no disagreement in the original position. If the veil of ignorance is well-motivated then it produces a single set of principles by itself. But there is another way to approach the problem that has a long history in the social contract tradition: appeals to social choice theory. Rousseau was one of the first political philosophers to grapple with the question of selecting among different political options. His discussion of voting rules is well known. Reflection on the discussion will help sharpen our own. Recall Rousseau's claim that,

¹¹ Hobbes 1994, pp. 76-77; Locke 2003, p. 160; Rousseau 1997, pp. 90-92.

There is only one law which by its nature requires unanimous consent. This is the social pact: for the civil association is the most voluntary act in the world; every man being born free and master of himself, no one may on any pretext whatsoever subject him without his consent. To decide that the son of a slave is born a slave is to decide that he is not born a man.¹²

Here Rousseau points out that a rule of unanimous consent is required for the *most general* form of political association, specifically, the form of civil association itself. If the social compact itself is not justified to a citizen, then she is not free, but a slave. However, once the social compact is made, it can easily accommodate disagreement: “Except for this primitive contract, the vote of the majority always obligates all the rest.”¹³ For the majority voting rule is the method by which members of the public delineate the general will and select among potential options for political issues as they arise. In the same discussion, Rousseau entertains a further question: “Yet the question is raised how a man can be both free and forced to conform to wills which are not his own. How are the opponents both free and subject to laws to which they have not consented?”¹⁴ In other words, how can it be that a man retains his freedom despite losing out on a vote on an issue of crucial importance to him? Rousseau replies that “the citizen consents to all the laws, even to those passed in spite of him, and even to those that punish him when he dares to violate anyone one of them. The constant will of all the members of the State is the general will; it is through it that they are citizens and free.”¹⁵ In other words, once a citizen is party to the social contract, the outcome of the procedure implied by the contract (the majority voting rule) is authorized by the general will and thus the will of each person as part of a corporate body. Citizens retain their freedom because the social compact is conclusively justified to them as is the decision rule implied by it. It is for this reason that Rousseau makes his oft mocked remark that upon losing a vote, “If my particular opinion had prevailed, I would have done something other than what I had willed, and it is then that I would not have been free.”¹⁶ In other words, if my “particular will” prevailed to become policy despite gaining a minority of the vote, I would have been *less free* than when the general will spoke against me.

¹² Rousseau 1997, p. 123.

¹³ Ibid., p. 124.

¹⁴ Ibid..

¹⁵ Ibid.

¹⁶ Ibid..

The convergence theorist follows Rousseau. There are many potentially justified principles, such as the three sufficientarian principles advanced by the factions of Alpha Centauri. But if a majority voting rule is publicly justified to all of them, they can select among any number of options and *be bound by the outcome by their very own wills*. This is so even if, say, Cyril's faction loses the vote, because in following the outcome of the vote they obey only themselves. Convergence reasoning enters the picture in this example when we ask why each faction regards the majority voting rule as fair. For the convergence theorist it doesn't matter why they regard it as fair so long as they have some conclusive reason *of their own* to regard the majoritarian decision procedure to carry the day. So the decision procedure can be justified by means of convergence reasoning but still yield determinate political principles by selecting among different principles. Given the pedigree of this response there is no reason to conclude that strategy (b) fails. I submit that the convergence theorist can draw on the social contract tradition to give the Rawlsian an adequate reply.

So far we have no reason to abandon the convergence standard. Convergence still provides methods of determining which political principles a society should follow. Convergence theorists can allow all public political reasons to reveal that certain political principles cannot be justified to all, but they can also appeal, following the great social contract theorists, to decision procedures justified according to the distinct, diverse reasons of each person. It seems that the pro tanto stage of justification should be abandoned. Convergence expresses two fundamental values of public reason liberalism: respect for reasonable pluralism and the protection of individual liberty. If the pro tanto stage is unnecessary, then given their own theoretical values, public reason liberals have no reason to employ it.

III: Eberle and the Agapic Pacifist

Rawls's combination of consensus and convergence is motivated by the fear that a convergence conception of public justification will yield indeterminate results. Political theory will turn up too many potential political principles and leave citizens with no rational method of selecting among them. Christopher Eberle has advanced the opposite objection: convergence justifications threaten to leave citizens without *any* principles or proposals that can be justified to all. In a recent article, Eberle points out that convergence promises to be far friendlier to citizens of faith than consensus views. Eberle writes, "[I]f a citizen or public official has what she regards as decisive religious objection to some coercive measure, then that coercive measure cannot be justified to her, in which

case it would be disrespectful ... to impose it on her.”¹⁷ Religious reasons can play a role in public justification because the convergence theorist permits them to undermine the justifications for particular coercive proposals. Accordingly, convergence “accords to religious citizens a *potentially decisive role* in determining the legitimacy of state coercion.”¹⁸ Conversely, secular citizens can undermine the justifiability of coercion as well, for if a secular citizen has a decisive secular objection to some coercive measure, then it is similarly defeated. Eberle thinks convergence enables secular and religious citizens (or differing groups generally) to undermine all reasonable proposals with respect to some key issue. This is the troubling dark side of convergence. If we permit a diversity of reasons, just as in strategy (a) above, the number of points of overlapping consensus may decrease dramatically. We allow more reasons to endorse *and* more reasons to reject. In contrast to consensus justifications, convergence permits diverse reasons to play both a constructive and destructive role with regard to the public justification of coercion. For Eberle, convergence “retains an exceedingly demanding conception of what makes for justified coercion”; on convergence, even coercion required for liberal democracy itself may not be justified.¹⁹ And so, Eberle concludes, it is “better to jettison the conception of justified coercion” than “to deny the legitimacy of those necessary measures.”²⁰

Eberle illustrates his point with the role that an “Agapic Pacifist” plays within public reasoning about national foreign policy. An Agapic Pacifist is one who thinks that Jesus’ command to Love Thy Neighbor bars the lethal use of violence; on Christian grounds, she opposes war. The Agapic Pacifist takes herself to have “compelling theological reasons” to reject the coercion required to protect rights. If we allow a great diversity of reasons into public justification, the Agapic Pacifist can threaten the legitimacy of “any and all wars.”²¹ Eberle concludes that “liberalism cannot survive the convergence conception of justified coercion.”²² Let us assume for the sake of discussion that the Agapic Pacifist is reasonable and rational, that is, she is generally willing to compromise with others (though not in this case, given what is at stake) and that she has coherent and sound epistemic commitments. The Agapic Pacifist threatens to make liberalism impossible because she defeats the justification for war, which is sometimes essential to maintaining a stable social order.

¹⁷ Emphasis mine. Eberle 2011, p. 12.

¹⁸ Ibid.

¹⁹ Ibid.

²⁰ Ibid., p. 13.

²¹ Ibid., p. 14.

²² Ibid.

Perhaps the convergence theorist could offer the Agapic Pacifist an exemption in response to her complaint: not only will she not be required to serve in any war, she will not be forced to finance any war. After all, the liberal state has long practiced a policy of religious accommodation, particularly with respect to conscientious objection. Eberle contends that this reply fails to give “due weight to the Agapic Pacifist’s concerns.”²³ The Agapic pacifist objects not merely to her killing anyone but to *anyone* killing anyone. Her government cannot be allowed to kill the innocent insofar as it is her government. Eberle claims that it is reasonable for the Agapic Pacifist “to insist ... that [her] government’s agents may not kill.”²⁴ Second, Eberle reminds us that the Agapic Pacifist is objecting to non-trivial policy, policy that might massively affect the Agapic Pacifist’s life. Thus, no matter the accommodation, “the liberal state’s employment of lethal violence will unavoidably and powerfully affect [her] well-being.”²⁵

Eberle’s response can be taken in two ways depending upon why it matters that the Agapic Pacifist’s well-being is affected. (1) On one interpretation, publicly justifying war to the Agapic Pacifist may be required simply because her well-being is affected. (2) Alternatively, publicly justifying war could be required because some principle of concern for well-being is already publicly justified to the Agapic Pacifist and her political community. The first interpretation holds that the fact that the Agapic Pacifist’s well-being is affected is a sufficient reason to require that a rule be publicly justified to her; the second interpretation holds that justification is owed to the Agapic Pacifist on the basis of already publicly justified rules against diminishing the well-being of others.

The first objection seems to fail, for a complaint of a loss of well-being is insufficient to trigger the requirement of public justification. Public reason liberalism is rooted in a presumption in favor of liberty that can only be met by a public justification. Thus, free and equal individuals have a duty to justify interfering with others. But why not think that the threat of a loss of well-being requires justification? It may initially appear that if Reba’s actions cause John to lose well-being that she has a complaint against him. Reba may seem presumptively obligated to avoid causing John to lose well-being. We can understand this reply as suggesting two different modifications of the liberty

²³ Ibid., p. 24, ft. 28.

²⁴ Ibid. We might deny the assumption that the Agapic Pacifist is reasonable to demand that her government not kill but I shall not pursue this line here.

²⁵ Ibid. Also see Horton 2010. While Eberle phrases his objections in terms of well-being, I think it is better understood as a point about membership in a state and what sorts of treatment state membership entitles one to demand. Can the Agapic Pacifist, qua member of her state, rightly demand that her state not engage in actions she regards as deeply immoral? This question is powerful independently of the question of whether the Agapic Pacifist’s well-being is affected.

principle: (i) count Reba's causing John to lose some well-being as a form of interference, (ii) replace the liberal principle with a presumption against a loss of well-being.²⁶

Reply (i) is implausible because if Reba causes John to lose well-being, he has not necessarily interfered with her. To give one example, if John adores Reba's character on his favorite TV show, and Reba decides to quit the show, John may well lose well-being in the process. Nonetheless, Reba does not interfere with John by quitting. Reply (ii) is implausible because it is excessively demanding. To use the same example, if John will lose well-being if Reba quits his favorite show, then given a presumption against causing losses of well-being, Reba must justify her change of occupation in terms he can rationally endorse.²⁷ In this case, then, John potentially has a veto over whether Reba can quit her job. This seems wrong: it places an enormous burden on Reba. John has the power to trap Reba into a job she dislikes. Further, John may well impose upon Reba's well-being in demanding that she not quit her job, which would trigger a similar requirement of justification. The idea of such a duty of justification on Reba's part (and John's) runs counter to our commonsense conception of the behavior we owe to each other.²⁸

The second version of Eberle's objection suggests that publicly justifying war is required by a publicly justified principle of concern for well-being. The objection probably shows that the Agapic Pacifist has a complaint against her government. However, public reason liberalism has the conceptual resources to answer the complaint. Any publicly justified polity will have a number of basic, widely justified political norms. In liberal democratic societies, these often include principles of free speech, freedom of religion, due process, the right to vote and the right to hold personal property. Further, there are typically dozens of standards that together represent a general presumption against harm. That is, liberal democratic societies tend to embrace a political harm principle: citizens have a claim against being harmed without justification. In cases where harm is imposed, governments are often required to provide compensation. If a harm principle is publicly justified in the Agapic Pacifist's society (let us presume that it will be), then the Agapic Pacifist may well have a complaint should her society go to war. But now that the challenge has been brought into focus, as it is no longer clear that a publicly justified polity is unable to respond appropriately. The Agapic Pacifist's complaint is essentially that she has been forced to pay for a *negative externality*

²⁶ A third option would be to embrace both presumptions, but I take it that this option is sufficiently strange and complex that it can be set aside.

²⁷ To simplify the example, assume that Reba will not lose significant well-being by ending the relationship with John.

²⁸ In conversation, Eberle has pointed out that this case can be dismissed because the concept of well-being in his objection need not be so subjectivist as to count canceling one's TV show as a major loss of well-being. A more morally significant loss might trigger a requirement of justification. This option is worth exploring. I will do so in the future.

produced by her state; the state has imposed a welfare loss upon her without justification. As a cooperative member of her society, the Agapic Pacifist objects to war and maintains that not forcing her to fight or pay for others to fight will not offset the potential costs to her well-being. This is a valid objection, but it is far from clear that there is no way to off-set the costs. Various forms of compensation might be made available, and it would seem to be unreasonable for the Agapic Pacifist to rule out as unjust any form of compensation that the state could offer her in good faith.²⁹

The power in Eberle's criticism is that convergence may initially appear to so substantially reduce the opportunities for justified coercion that a society cannot morally perform its most basic functions. But Eberle's case of the Agapic Pacifist fails to undermine the case for convergence. Sometimes convergence will reduce opportunities for justified coercion, but in other cases, convergence may increase (perhaps significantly) the number of eligible coercive proposals. Thus, it is unclear in the abstract whether convergence will produce unmanageable indeterminacy or a debilitating inability to coerce. Eberle's argument generalizes from a single case to determine the effect of convergence upon the vast network of rules that comprise the social order, yet a single case cannot do the job. Any publicly justified decision must occur with reference to a network of established laws and norms on either a consensus or a convergence view. The question of convergence is merely the question of whether private reasons may be used to thread the network.

IV: Macedo and Common Rationales

The first two objections maintained that convergence renders political justification impossible; the third objection argues that while convergence is a viable form of public justification, it undermines some of public reason's other foundational commitments.³⁰ Stephen Macedo has complained that convergence's rejection of the necessity of shared commitments entails abandoning public reason's traditional ideal of a shared commitment to justice. Macedo criticizes the convergence view of public reason on the grounds that it "has no special relation to the constitution of the public sphere."³¹

²⁹ One might worry here that public reason liberalism cannot justify war anyway, as it cannot justify the coercion required for war to an opposing army. However, many norms that we need can be justified within the framework of public reason. I assume that many norms required for war are of this sort.

³⁰ Jonathan Quong has a detailed argument against convergence on the grounds that it violates a plausible principle of sincerity. However, we have already shown that the principle of sincerity is determined after the standard of justificatory reasons is determined, so that sincerity cannot be incompatible with convergence (Ch.6, VII). For Quong's objection, see Quong 2010, pp. 265-273.

³¹ Macedo 2011, p. 3.

Public reason is supposed to “empower a common point of view.”³² Those who share the convergence view, including Gaus and Jeffrey Stout, “downgrade or reject the ways in which liberal public reason is concerned to justify shared principles on the basis of *common knowledge*.”³³ Instead, they embrace “privatism” or “partialism,” placing an excessive emphasis on reasonable pluralism. Consequently, these privatistic, partialist views generate an unacceptably thin conception of politics. A thin conception of politics is one that encourages citizens to find their sense of community only in “the partial associations of civil society, and the bosom of family and friends.”³⁴

Macedo recognizes that some may find a thin conception of politics unobjectionable and even attractive. But a thin politics has a severe disadvantage in comparison with a thicker politics that relies largely on shared values: privatism enervates the public sphere’s ability to share common knowledge. Common knowledge is not merely information shared within a group but requires knowing about what others know. To better understand Macedo’s conception of common knowledge, we can compare it to Rawls’s arguably quite similar conception of publicity. To achieve the Rawlsian ideal of publicity, a society must recognize that it affirms common principles of justice. Rawls distinguishes three levels of publicity. The first level “is achieved when society is effectively regulated by public principles of justice: citizens accept and know that others likewise accept those principles, and this in turn is publicly recognized.”³⁵ Everyone understands that a society’s basic structure is justified in terms of commonly accepted practices of inquiry and reasoning. The second level of publicity requires that citizens of a well-ordered society have shared beliefs “in the light of which first principles of justice themselves can be accepted, that is, the general beliefs about human nature and the way political social institutions generally work, and indeed all such beliefs relevant to political justice.”³⁶ A society must share beliefs about human nature and the functioning of social institutions in order to have a common framework for cooperation and moral justification. The third level of publicity concerns what Rawls calls “the full justification of the public concept of justice as it would be presented in its own terms”; this level must include all of the considerations we bring to the table when generating a conception of justice in the first place.³⁷ The full justification—or, as we called it, the convergence stage—must be “publicly known, or better, at least [known to be] publicly

³² Ibid., p. 1.

³³ Ibid.

³⁴ Ibid., p. 16.

³⁵ Rawls 2005, p. 66.

³⁶ Ibid., p. 67.

³⁷ Ibid.

available.”³⁸ All members of the public must either know or be able to know that full justification has occurred, that each citizen has accepted the political conception. When a well-ordered society has all three types of publicity, Rawls claims that it meets “the full publicity condition.”³⁹

Like common knowledge, full publicity is important because it is appropriate to our status as “reasonable and rational citizens who are free and equal.”⁴⁰ In a society with full publicity, the exchange of reasons will render citizens “confident that this avowed reckoning itself will strengthen and not weaken public understanding.”⁴¹ As we saw above (Ch.5, III.ii), Micah Schwartzman has recently argued that full publicity has value for a number of reasons. For instance, it may be “necessary for democratic accountability” and could “enhance the quality of political decisions.”⁴² Full publicity will also make justifications for coercion publicly available so they can be evaluated and examined.⁴³ Similarly, Macedo suggests that common knowledge “helps people coordinate their actions in service of mutual advantage” and aids “cooperation on the basis of shared values,” both benefits akin to those Rawls ascribes to full publicity.⁴⁴ When citizens speak in shared terms, their arguments can be easily evaluated by their interlocutors because they share evaluative standards and reasons with us. Information is more easily shared when it is expressed in a common political language.

According to Macedo, convergence lacks substantive common standards of reasoning. Consequently, a society rooted in the privatistic conception of public reasoning fails to promote common knowledge:

But our capacity to assure one another of our shared commitment would seem more solidly grounded and stable, and also more mutually intelligible, if we share a rationale, or agree on a widely convincing proof. Then we not only agree each for our own reasons, but share a powerful rationale in common.⁴⁵

Mutually intelligible shared commitments provide benefits that privatistic societies fail to generate in adequate supply. For Macedo, shared commitments are similar to public goods; without a civic

³⁸ Ibid. Rawls allows for the weaker, “at least” condition because many may not want to expend the cognitive resources required to acquire the relevant knowledge. For more on Rawls’s various conceptions of publicity, see Freeman 2007.

³⁹ For a more extended discussion of publicity, see Ch.3, III.iii.

⁴⁰ Rawls 2005, p. 67.

⁴¹ Ibid., p. 68.

⁴² Schwartzman 2011, pp. 7-8.

⁴³ Ibid., p. 9.

⁴⁴ Macedo 2011, p. 19.

⁴⁵ Ibid., p. 16.

guarantee that reasoning will occur on the basis of shared commitments, citizens will go their own privatistic way and the norm will evaporate. Macedo marshals a key example to illustrate his point. Privatistic versions of public reason will tend to disadvantage the weak and marginalized. They are “thin” because they contain few shared norms that members of society can be called to follow. These norms often protect the weaker members of society. In a society with a thin conception of public reason, the “relatively powerful” will have no problem, but the powerless will fare poorly. For “who will have the resources to engage in private self-help and defense when others act unfairly?”⁴⁶ The poor and marginalized depend on shared commitments to protect themselves against the powerful. On convergence, more powerful citizens can defect on common moral norms that protect the poor and marginalized, leading the norms to collapse or weaken. To put it another way, norms are upheld by shared beliefs and practices about what is right and good. If the powerful do not follow these norms in order to secure advantages for themselves, they undermine the joint basis of enforcement of these norms.

To what extent does convergence create problems for common knowledge and full publicity?⁴⁷ It is not clear from Macedo’s argument. Convergence doesn’t rule out common reasoning; it only allows diverse reasoning. To refute convergence, Macedo must argue that broadening permissible justifications leads to a paucity of shared norms needed to protect potentially marginalized citizens, but Macedo makes no such argument. For all we know, convergence reasoning may increase the opportunities to develop shared commitments and values. Convergence permits many forms of overlap that citizens could not form under consensus. Because convergence is less restrictive, citizens may be able to converge on common proposals and norms through a wider range of mechanisms and reasoning systems. On the whole, convergence increases the amount of information available to the public. With more reasons on the table, citizens become aware of a broader range of considerations and have more resources from which to work out shared political principles and institutions. Unless the increase in information creates some sort of “noise” effect, where opportunities for public justification are obscured, it is hard to see why convergence should pose a problem for creating common commitments.

⁴⁶ Ibid.

⁴⁷ Recall that we have already addressed publicity concerns in Ch.3. There I argued that the conception of social rules justified in second-personal terms is sufficient to satisfy concerns about publicity (Ch.3, III.iii).

One might worry that by permitting private reasons to defeat shared commitments and proposals, convergence undermines opportunities to develop them.⁴⁸ If so, convergence will take many options off the table relative to symmetric and consensus standards; what convergence gives in extra information and opportunities for overlap, it might take away. But as we saw at the end of Section III, it is unclear whether convergence will unacceptably add or subtract from the number of coercive laws that can be justified. Convergence will increase opportunities for coercion with respect to some matters and reduce them with respect to others.

That said, the possibility of defeat is not a complaint against convergence, for without the ability to defeat unjustified coercion, laws and policies passed will often be unjust. On the convergence view, a conclusive private reason to reject a shared rationale or proposal will similarly show that the commitment or proposal on the table is unjustified.⁴⁹ When the airing of unshared, inaccessible reasons removes opportunities for public justification, it does so by revealing citizens' true reasons. Consensus obscures private reasons, preventing them from playing a justificatory role. If the ideal of public reason is to create a political society whose coercive institutions are justified, then citizens must be rationally committed to the coercion imposed upon them. Thus, citizens should be rationally committed by their entire rational will, rather than only the "public" part of their reasoning.⁵⁰ Consensus theorists fail to realize that by restricting the set of reasons relevant to public justification they give up the ideal of justification to the full reason of the individual, which is the only way to preserve her dignity as a free and rational being.

Further, note that even Rawlsians reject political conceptions that cannot achieve an overlapping consensus. A political conception is only justified if it is *fully* justified or shown to fit within all reasonable comprehensive doctrines in a well-ordered society. Consequently, the Rawlsian view *cannot possibly* generate a thicker conception of politics—one with more norms—than a mere convergence view. The set of publicly justified principles on convergence is a superset of the principles recognized by Rawlsians. The only way for convergence to generate more opportunities for defeat is if it is combined with a fine-grained individuation of coercive actions. Rawls believes that only constitutional essentials must be publicly justified: "on matters of constitutional essentials

⁴⁸ It is worth noting that the extent to which convergence allows in more defeater reasons than consensus depends on whether a symmetry requirement is added to either standard. Typically consensus is tied to symmetry and convergence to asymmetry, so the argument of the piece simply assumes that the two options available are symmetric consensus and asymmetric convergence.

⁴⁹ This is one reason that Quong substantially restricts the role of the overlapping consensus in his account of public reason. Quong 2010, p. 336.

⁵⁰ For the original articulation of this ideal, see Rousseau 1997, p. 50.

and basic justice, the basic structure and its public policies are to be justified to all citizens.”⁵¹ But if convergence requires that each law be publicly justified, then perhaps more opportunities for defeat will present themselves. That said, the set of coercive actions is an independent variable within the Public Justification Principle. If there are good reasons to finely individuate coercive actions on convergence, these same reasons should provide grounds to finely individual coercion on the Rawlsian dual conception. There is no obvious reason to treat the convergence and Rawlsian views differently.⁵²

Finally, convergence can defend the poor and marginalized. Consensus reasoning restricts public reason to shared and accessible reasons and values. By allowing private reasons into public justification, convergence permits considerations to enter public discourse developed by *poor and marginalized communities*. Consider the African-American community. Due to its shared history and experience, blacks developed several languages of empowerment that many whites had trouble understanding. From shared linguistic conventions, to shared historical documents, to shared interpretations of common values, the black community (or communities) developed traditions of reasoning that are more appropriately understood as “privatistic.” This is true of the Black Power movement which latched onto a necessarily private value in order to organize African-Americans. It is hard to square black communities’ use of private traditions of reasoning with a strong commitment to consensus reasoning. I submit that the same holds for the feminist community; in fact, many feminists have argued that women tend to appeal to different ethical concepts than men do and accordingly speak their own language of “the ethics of care.”⁵³ While care ethics need not be restricted to women, a convergence theorist can argue that as a matter of historical fact, care ethics has been a source of reasoning of particular import to women and has thereby acquired a non-public element.⁵⁴ Consensus may help protect the poor and marginalized but in the end, consensus approaches threaten to close off many public paths to genuine protection of their interests. Convergence is a more active and empowering form of public reason because it is friendly to the private reasoning traditions of marginalized groups.

V: The Consequences of Convergence

⁵¹ Rawls 2005, p. 224.

⁵² For instance, Quong thinks that even on the Rawlsian pictures, public justification should apply to laws. See Quong 2004.

⁵³ For one prominent example of such an argument, Noddings 2003.

⁵⁴ Rawls 2005, p. 37.

The foregoing arguments, in my view, conclusively show that convergence is a viable conception of public justification. When combined with the arguments of Chapter 6, the arguments of Chapter 7 vindicate convergence. Chapter 6 provides reason to think that convergence is superior to consensus standards and Chapter 7 provides reason to think that convergence imposes few first and second-personal costs and suffers from no obvious defects. We may now draw out the consequences of convergence for public reason liberalism.

V.i Three Benefits of Convergence

Convergence advances the dissertation's project in two ways. First, it eliminates the need for restraint. To recall, a principle of restraint is a norm that requires citizens not to offer or act upon religious reasons in public political life (Ch.2, II.iii). Since restraint is the primary integrity-threatening norm recommended by public reason liberals, defeating the case for restraint helps undermine the conflict between public reason and integralist liberals. Without restraint, integralist objections to public reason liberalism are defused. The proposal-restraint implied by convergence applies to few citizens and only restricts the reasoning of Supreme Court Justices, and related adjudicative bodies (Ch.6, VII). Second, by eliminating restraint, convergence preserves integrity and respect for reasonable pluralism. Not only does convergence impose few integrity costs, but it expresses respect for a broader range of positions than the standard conception by permitting diverse reasons to enter the public square. In doing so, convergence extends and unites social morality and individual ideal.

V.ii Convergence and the Future of Public Reason Liberalism

Another consequence of convergence cannot be ignored, for adopting convergence promises to alter the public reason tradition. I believe that the objections reviewed in Sections II-IV all stem from the implicit observation that convergence makes public justification harder and more complicated. Allowing a vast array of private reasons into public justification threatens an unmanageable increase of information. Rawls tried to tame this unmanageability by creating a theoretical bottleneck that restricts the reasons that can enter into political theory and practice. Eberle sees the new information as a normative hailstorm, destroying many fruitful opportunities for social cooperation. Macedo worries that the flood of information unleashed by convergence will

decimate shared norms and commitments. A smaller pool of reasons with similar characteristics may allow a more stable social order to form. A consensus standard structures reasons in a more static and seemingly solid social edifice; convergence threatens shared institutions with reasons that could undermine them. Without these common structures, many valuable public institutions could collapse. The objections can be summed up as follows: *convergence presents public reason with information overload.*

There are two relevant information problems. The first holds that convergence makes it harder for citizens to determine how to conduct public dialogue. With the smaller set of symmetric, shared and accessible reasons provided by consensus, it is easier to discern how to treat others with discursive respect: simply offer shared or accessible reasons in argument. Convergence standards imply fewer clear guidelines for good citizenship. Consequently, it may prove difficult for citizens to form shared commitments and rationales. Thus excessive information obscures how to rightly engage in *the practice of citizenship*. Recall from Sections II and III that Eberle and Rawls cite opposing problems for convergence. Consequently, the second information problem suggests that convergence makes less clear which political principles and laws are justified, as political theory alone cannot predict whether convergence poses the opposing problems leveled against it by Rawls and Eberle. The political world of convergence is less determinate, less predictable and more dynamic, and so excessive information obscures how to rightly formulate *the theory of justice*.

These two challenges should excite public reason liberals, not intimidate them, as they both present the opportunity for theoretical development. The first information problem can be met by showing that convergence requires little of citizens and pointing to the resulting benefits, as we have seen above.⁵⁵ The second information problem suggests that abstract political theory can tell us less about which norms are justified. Given that political theory cannot determine which reasons are permissible, the door is open for social processes to help societies arrive at better norms through a free exchange of ideas. While sensitivity to such processes is part of the deliberative democracy literature, convergence theorists must move beyond deliberation in order to solve the problem of public justification.⁵⁶ As we shall see in Chapter 10, the information problems will likely prove

⁵⁵ Citizens should be honest in offering their best reasons, of course (save in obvious strategic situations). They should also, at least for pragmatic reasons, offer arguments they think will be compelling to their interlocutors (again, when they are not in strategic situations). But being honest and practical is far less burdensome than the principles of restraint embraced by consensus liberals.

⁵⁶ Jurgen Habermas has made perhaps the most renowned attempt to integrate deliberative democracy with a theory of social cognition. See Habermas 1999, esp pp. 116-170.

unmanageable at the level of explicit, conscious cognition. Instead, public reason liberalism must recognize the many, diverse interactions generated by convergence justifications. Citizens will not be able to discern which reasons are most relevant to public justification as they will be too numerous and will enter into public justification in too many ways. Therefore, a turn towards social cognition may eclipse the central importance of deliberation to the political process. Public reason liberals could instead focus on determining how political institutions can track the development of norms and reasons that arise from the social practices and processes of different societies. Publicly justified institutions will help to register the views of citizens as inputs and output publicly justified laws and policies. Thus public reason under convergence takes an institutional turn.

The institutional and social epistemological aspects of convergence also suggest a shift of importance within the public reason tradition from political philosophers to social scientists. If political theory alone cannot determine which norms are justified, then the political philosopher's job becomes more circumscribed, and if achieving public justification requires tracking complex forms of social cognition that do not always involve explicit, deliberative reasoning then the skill set of the philosopher—dialogue, deliberation and criticism—may not effectively track the reasoning of citizens. That said, while convergence suggests that the political philosopher is less central to the development of just institutions than consensus liberals have held, the shift can be understood in more constructive terms. If political philosophers have reason to further emphasize institutional design and social cognition, perhaps they should give more attention to those political philosophers who were themselves great social scientists, like Adam Smith, and to the largely ignored political economy of philosophers whose other work receives great attention, such as David Hume, John Stuart Mill and Henry Sidgwick.

V.iii A Profound Weakness for Convergence

Despite these benefits, convergence cannot perform the work of reconciliation alone. To see why, let's recall the Public Justification Principle:

The Public Justification Principle: A coercive action C is justified iff every member of the public P has (a) conclusive reason(s) R to endorse C.

Chapters 5, 6 and 7 only analyze reasons R. But the success of convergence depends on how we idealize ordinary citizens, that is, how we fill out the members of the public P variable. The problem is that we might idealize radically. In other words (as we shall see in Chapter 8), we might require that members of the public be fully rational and fully informed. And if they are so idealized, their reasons could be radically restructured. Fully informed and rational agents may pursue completely different projects and have different values than their counterparts. Accordingly, if citizens are radically idealized, they may be permissibly coerced on the basis of reasons acknowledged only by their idealized counterparts. In this way, radical idealization might obliterate a citizen's present projects and plans. If so, radical idealization could impose severe integrity costs. What's worse, radical idealization threatens to impose second-personal costs as well. Many public reason liberals engage in radical idealization in order to render the reasons of citizens identical. This is done to make the problem of political justification more tractable. In doing so, public reason liberals undermine reasonable pluralism at the stages of political justification involving radical idealization. In this way, the diverse reasoning and reasons of citizens are ignored. Even with convergence, public reason liberalism may still impose considerable first and second-personal costs, thereby undermining our entire project of reconciliation. Consequently, we must push back against radical idealization while avoiding the dangers of populist conceptions of idealization (Ch.1, IV.ii). Chapter 8 sets out the conception of idealization employed by mainstream public reason liberals. Chapter 9 argues against the standard conception. Chapter 10 will defend a conception of *moderate* idealization that, when coupled with convergence, can reconcile religious conviction and the demands of public justification, the first and second-personal points of view. We shall see that citizens of faith will retain their freedom within the bounds of public reason.

Chapter 8: Conceptions of Idealization

Resolving the tension between the first and second-personal elements within liberal political theory requires developing a conception of public reason that substantially reduces costs to integrity while expressing public reason's foundational values. Chapters 5, 6 and 7 completed the first half of this project by developing a convergence conception of public reasons. I argued that convergence is superior to consensus conceptions on both the first and second-personal metrics explored in Part I. However, convergence alone cannot resolve the tension between the first and second-person. Despite eliminating the motivation for principles of restraint, convergence conceptions are compatible with views of public reason that model members of the public apart from their core projects and plans. Public reason liberalism cannot represent the reasoning of members of the public as having no knowledge of their identities and projects. Thus, even if public reason liberals embrace convergence, their conception of public reason can still violate integrity because first-personal reasons will not enter into public justification. Accordingly, they will be unable to block coercion that could stifle individual liberty, including religious liberty. To give one example, Rawls's veil of ignorance idealization attributes reasons to citizens divorced from their conceptions of the good. It is for this reason that his version of public reason liberalism has raised the hackles of many philosophers on the grounds that it abstracts excessively from the integrity of citizens. This criticism is not limited to integralists, but comes also from communitarians like Michael Sandel and other contractualist liberals like Brian Barry.¹

That said, public reason liberals have good reason to idealize (Ch. 1, IV.ii). The ideal of public reason is rational justification of coercion to all; when such justification is achieved, the freedom and equality of citizens is preserved despite the necessity of coercion for social life. If rational justification is intelligible justification, then public justification must be proceed based upon reasons with good epistemic credentials. Without idealization, public justification will be held captive to errors in reasoning and faulty information; consequently, citizens will be coerced on the basis of poor reasons. Idealization is a required part of public reason.

Public reason liberals understand the risk of insufficient idealization but have rarely recognized the possibility of erring in the opposite direction. Yet radical idealization has costs. First, it threatens to disconnect the idealized parties from the citizens they're supposed to model. Citizens

¹ Barry 1996, pp. 57-61; Sandel 1982, pp. 161-5.

will have no reason to regard the decisions of the parties to a deliberative situation as binding if the deliberators do not resemble them. If public reason liberals disconnect their models of citizen reasoning from citizens' presently recognized reasons, the models lose plausibility. Second, radical idealization threatens to destroy reasonable pluralism. It does so in large part because public reason liberals assume that the more we idealize, the more similar we become. For Rawls, reasonable pluralism does not exist at the level of the parties to the original position; reasonable pluralism enters public reason at the *second* stage of political justification, not the first. Despite the fact that the parties know that reasonable pluralism exists and that they represent people who do not share the same conception of the good, their reasoning is still identical. By idealizing so radically, Rawls's original position raises the ire of his integralist critics perhaps as much as his principle of restraint, as the original position seems to generate principles of justice based on the reasoning of "disembodied selves".²

Just as Chapter 5 outlined the standard conceptions of public reasons, the present chapter outlines standard conceptions of idealization. It also details the costs and benefits of the standard approach. In short, this chapter maps the conceptual territory of idealization within public reason. To explain, recall:

The Public Justification Principle: A coercive action C is justified if and only if each and every member of the public P, has (a) conclusive reason(s) R to endorse C.

Chapters 8, 9 and 10 switch focus to the members of the public P from reasons R, in particular I consider the various dimensions of idealization of the members of the public—reasonableness, rationality, information and coherence. This chapter proceeds in four sections. Section I outlines the conceptions of idealization endorsed by Rawls, David Gauthier and Jürgen Habermas, who arguably have the most developed views on these matters. In Section II, I draw on these theories to describe the rationality, information and coherence dimensions of idealization. Section III employs the three theories to explain the concept of the reasonable within public reason liberalism and its role in idealization. Section IV concludes by discussing what's at stake in a theory of idealization.

I: Standard Models of Idealization: Rawls, Habermas, Gauthier

² The disembodied selves charge is advanced not only by Sandel, but by Charles Taylor. See Taylor 1989, p. 89.

I.i What is Idealization?

Public reason liberals must have a theory of justificatory reasons. In other words, they must have an explanation for which reasons citizens have in order to show that citizens have a conclusive *reason* to embrace coercion. Convergence conceptions permit the entire class of public reasons to enter into public justification. But remember that public reasons are all those reasons bearing on the justification of coercion that are *intelligible* to citizens. In other words, these reasons are ones that members of the public believe can be properly inferred from a citizen's belief-value set. Idealization explains what is to count as proper inference. To engage in *radical* idealization is to attribute reasons to citizens on the basis of inferences from one's belief-value set of a very strong sort. I believe that idealization is best understood as having four dimensions: rationality, information, coherence and reasonableness. Inferences are sound on radical idealization views when they are made with (i) complete or perfect information, (ii) perfectly sound inference rules and (iii) wholly coherent with all other sound inferences from a citizen's belief-value set and (iv) fully reasonable. Radical idealization ascribes reasons to a citizen that only very long, perfectly informed and (morally) reasonable inferences would lead the citizen to endorse.

In these remaining chapters, I will often speak of "models" of citizens. Within the public reason literature, a model of citizens is simply a representation of reasons that are openly justified to them. For instance, if a model of a citizen is perfectly rational then it will ascribe reasons to citizens based on perfect reasoning. The model is an interpretation of proper inference given open justification. Rawls's parties to the original position, for instance, provide an interpretation of proper inference for citizens that lead them to endorse reasoning characterized by the veil of ignorance. Idealization is not an ideal to which citizens should aspire, nor does it model particular features of citizens' moral psychology. Instead, it is an explanation and interpretation of their *reasons*. Thus idealization often involves a kind of abstraction from citizens' present belief-value sets in order to arrive at the reasons they are plausibly committed to. Idealization aims to explain why citizens have some reasons rather than others, which in turn is required to satisfy the Public Justification Principle. I will also discuss the crucial *relationship* between real-world agents and the idealized agents who model their reasons. The idealized agent is simply a representative model of a real-world agent's reasons. In other words, the model is a place-holder that represents an appropriately idealized belief-value set of a real-world agent or citizen. Idealized agents, therefore, are not choosing principles or

commitments for real-world agents. Instead, they are simply representations of updated and corrected belief-value sets of citizens.³

A critic could reply that the standard conception of idealization is typically employed makes my criticisms irrelevant. For instance, Rawls's idealized agents—the parties to the original position—select principles of justice for real-world citizens, or, at least, citizens of a well-ordered society. For Rawls, confusing the two agents is to make a serious mistake.⁴ In this way, idealized agents select principles for non-ideal agents. If so, then it appears that the models employed by public reason liberals are themselves ideal *advisors* and not agents that *represent* the reasoning of citizens. In reply, consider that within public reason liberalism, models of idealization determine the reasons to which citizens are committed such that when they are coerced on the basis of those reasons they have no complaint of injustice or mistreatment. Thus, in my terms, Rawlsian parties to the original position are not *themselves* ideal agents but are a part of an idealization model that contains Rawls's entire theoretical apparatus. The “idealization” is not merely the original position but the process of reason attribution of which it is a part. An agent idealized in the standard way is a representation of the reasons generated by a theoretical apparatus like Rawls's. For instance, suppose that Rawls's idealization shows that parties to the original position will embrace the difference principle. In my terms, agents idealized in Rawls's way are committed to recognizing the difference principle, but not because they themselves *are* parties to the original position but because they are the agents who will recognize the difference principle as contained within or inferable from their belief-value set. Thus when I criticize the standard conception of idealization, I criticize, say, the Rawlsian idealization process as a whole.

Finally, I must pre-empt an important challenge. A reasonable reader may deny my claim that idealization should provide citizens with reasons that help them solve real-world problems. She might argue that these real-world problems are irrelevant to how selecting principles of justice should proceed. Parties responsible for selecting principles of justice are not making real-world choices; after all, they're selecting principles of justice. But this would be to, again, misunderstand my conception of idealization. Idealization is simply a method of fleshing out the ideas of open justification and intelligibility that in turn specify the set of a citizen's public reasons (Ch. 4, II.ii). When we idealize an agent's rational capacities, we are only interpreting an agent's set of commitments in light of greater rationality, information and coherence. In my view, agents idealized

³ I will address an “ideal advisor” account of idealization below, where an idealized agent selects principles for a distinct unidealized real-world citizen, but this case is distinct.

⁴ Rawls 2005, p. 28.

in the standard way do not select principles of justice. Instead, idealized agents set the limits of justifiable norms in terms of the reasons they endorse. The Public Justification Principle determines which principles, norms, policies, etc. can be justified because it fleshes out an account of open justification and intelligibility that in turn delineate the set of citizens' reasons. Bona fide principles of justice are those compatible with the set of the public's conclusive reasons. To be sure, a model could be constructed to help determine which principles of justice those would be, but I am not attempting to build such a model. Instead, I construe the standard conception of idealization as an attempt to model citizens' reasons by providing an interpretation of open justification and intelligibility in terms of reasons that would be endorsed by fully rational, fully informed agents with perfectly coherent preferences and beliefs. I emphasize these matters because idealization is one of the wilderness areas within public reason, where many of its proponents and detractors become lost and confused.⁵

Let us now elucidate the standard conception of idealization by examining the conceptions of idealization defended by Rawls, David Gauthier and Jürgen Habermas.

I.i: Rawls

Rawls's conception of idealization is the most well-known, the most notorious and the most misunderstood, so it is best to begin with it. For Rawls, the aim of a theory of justice is to provide a justification for the claims that citizens make on each other's conduct. In order to legitimately obligate all citizens, principles of justice must be mutually acknowledged by "free persons who have no authority over one another."⁶ Rawls realizes, however, that gaining traction on the problem of justification is difficult, and responds by converting it into a social choice problem: "the question of justification [can be] settled by working out a problem of deliberation."⁷ Principles of justice are justified if they are selected by parties to a properly specified choice situation. The choice situation, however, must have certain properties, including aspects of "formal justice" such as the "impartial and consistent administration of laws and institutions."⁸ Thus, the "somewhat unusual conditions" of the original position serve to "make vivid to ourselves the restrictions that it seems reasonable to

⁵ Arguably a confusion on this point led to the liberalism-communitarianism debate as kicked off by Michael Sandel's criticism of *A Theory of Justice*.

⁶ Rawls 1958, p. 179.

⁷ Rawls 1971, p. 16.

⁸ *Ibid.*, p. 51. Sidgwick 1907, p. 267.

impose on arguments for principles of justice.”⁹ Idealization (altering the parties’ capacity for reasoning and information) is required to meet the conditions for choosing principles through constraints required by the ideals of formal and substantive justice. I should emphasize that Rawls’s parties to the original position are not *themselves* idealized versions of actual citizens; the original position “is set up by you and me” and thus “the nature of the parties is up to us: they are merely the artificial creatures inhabiting our device of representation.”¹⁰ As stated above, the parties are not meant to serve as models of human moral psychology.¹¹ In other words, the parties are not abstract representations of our psychologies so much as they are abstract models of our best *reasoning*. Deliberators represent the problem of justification rooted in our shared considered judgments about justice.

Rawls’s parties are idealized models because they are taken to be rational and reasonable. The reasons they generate will be those a rational and reasonable person would endorse. These reasons will be ascribed to actual citizens in turn. The parties are rational because they take “the most effective means to given ends.”¹² A rational person is said to “have a coherent set of preferences” and “ranks ... options” by how well they satisfy those preferences.¹³ The parties also know that it is public knowledge that they are capable of having a “sense of justice”, which is assumed to insure “the integrity of the agreement made.”¹⁴ The parties also have access to forms of information unavailable to normal human beings. For instance, they know and understand:

- (i) General facts about human society;
- (ii) Political affairs;
- (iii) Principles of economic theory;
- (iv) The basis of social organization;
- (v) The laws of human psychology;¹⁵

However, the parties also lack information. They are characterized as “reasonable” because they lack information that is “irrelevant from the standpoint of justice”; justice requires excluding “the

⁹ Rawls 1971, p. 16.

¹⁰ Rawls 2005, p. 28.

¹¹ *Ibid.*

¹² Rawls 1971, p. 12.

¹³ *Ibid.*, p. 124.

¹⁴ Rawls offers some special assumptions, though, including that “a rational individual does not suffer from envy”, which means that “he is not ready to accept a loss for himself if only others have less as well.” *Ibid.*, p. 125.

¹⁵ *Ibid.*, p. 119.

knowledge of those contingencies which sets men at odds and allows them to be guided by their prejudices.”¹⁶ The veil of ignorance serves to “nullify the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage.”¹⁷ Each party is deprived of the following information:

- (i) His place in society;
- (ii) His class position of social status;
- (iii) His natural assets, abilities, intelligence and strength;
- (iv) His conception of the good, the particulars of his rational plan of life;
- (v) The special features of his psychology (including, say, his degree of aversion to risk);
- (vi) The particular circumstances of his society and which generation he belongs to;¹⁸

Rawls recognizes that many will find the original position bizarrely abstract. Some will judge it irrational to choose principles of justice while being deprived of information.¹⁹ But this simplification is essential to insure that the choice situation is one where “each is convinced by the same arguments” and “unanimous agreement” is achieved; the parties are therefore not “bargaining in the usual sense.”²⁰ The reason abstraction is important for Rawls is that he believes it will harmonize the parties’ reasoning and lead them to converge on choosing a single set of principles of justice. In this way, the restrictions on information are crucial because “without them we would not be able to work out any definite theory of justice at all” but would rather be required to be content “with a vague formula.”²¹ As I emphasized earlier (Ch. 7, III), Rawls constructs his deliberative situation in order to avoid the problem of indeterminacy; and in fact, he was so determined to achieve determinacy that he ensured that the parties would agree unanimously on principles of justice (which is not to imply that unanimous agreement is the only way to ensure determinacy).

Our review of Rawls already reveals how he will fit into idealization’s four dimensions. The Rawlsian conceptions of rationality and coherence are radical: the parties cognize perfectly and have completely coherent preferences over various conceptions of justice. Rawls deprives parties of an enormous amount of particular information but provides the maximal degree of general

¹⁶ Ibid., pp. 16-17.

¹⁷ Ibid., p. 118.

¹⁸ Ibid., p. 118.

¹⁹ Ibid., p. 120.

²⁰ Ibid.

²¹ Ibid., p. 121.

information, and Rawls models reasonableness of some condition of fairness and reciprocity through the veil of ignorance.

I.ii: Gauthier

In *Morals by Agreement*, Gauthier argues that citizens require a social morality in order to secure the benefits of social cooperation. But for Gauthier, unlike Rawls, moral and political norms are justified by showing that instrumental rationality requires individuals to embrace a scheme of social rules. Individuals, “faced with the costs of natural or market interaction in the face of externalities agree to a different, cooperative mode of interaction”; rational individuals do so in order to “maximize their own utility.”²² Gauthier’s main concern is to show why anyone should accept the restraints morality places on individual behavior. His proposal, in short, is to show that even a very thin notion of rationality—instrumental rationality—can justify moral constraints to all; the reason that John can claim authority over Reba is because it is in Reba’s self-interest to allow John to do so. Gauthier’s conception of instrumental rationality is in many ways identical to the conception employed by standard interpretations of decision theory, where individuals are rational when they act on a preference ranking that meets the axioms of completeness and transitivity.²³ However, Gauthier argues that instrumental rationality requires agents to internalize a cooperative disposition, that is, to become “constrained maximizers” who cooperate with fellow cooperators but who defend themselves against exploitation. Constrained maximization is rational because “the disposition to make constrained choices ... is utility-maximizing.”²⁴ An agent acts morally when she acts on this cooperative disposition.

Like Rawls, Gauthier employs the idea of an impartial or impersonal point of view. But his impartial point of view, the “Archimedean Point”, is used to generate impartial rules from instrumental rationality. Only rules selected from this point can “possess the moral force needed to govern the moral realm.”²⁵ Since Gauthier’s theory is a true bargaining theory (unlike Rawls’s), bargainers recognize their distinct interests when they negotiate about how to split the gains from

²² Gauthier 1986, pp. 145-6. However, Gauthier does maintain that a satisfactory theory of rationality does not require rational agents to defect in prisoner’s dilemmas.

²³ Ibid. pp. 39-42. Gauthier, however, importantly alters the theory of instrumental rationality. There are also important alternatives to the standard interpretation. Gerald Gaus argues, for instance, that decision theory and instrumental rationality should not be identified. See Gaus 2011, pp. 53-100.

²⁴ Gauthier 1986, p. 183.

²⁵ Ibid., p. 233.

social cooperation. In order for rules to be impartial, they must be selected from a point of view abstracted from contingencies that would generate partiality. Idealization is required for agents to reason from this point; in response to the requirement, Gauthier develops a conception of an “ideal actor” which reasons from the Archimedean point.

The ideal actor is “fully rational.”²⁶ She is also “fully informed”, though not “about her own particular capacities and preferences, but about human capacities, preferences, and circumstances.”²⁷ The ideal agent possesses an “internally rational system” of preferences and beliefs.²⁸ Gauthier thus assumes that individuals reason perfectly and know all general information relevant to selecting principles of morality and justice, much like Rawls. The ideal actor also does not know her particular capacities and preferences, for reasons much like Rawls’s: “the appeal to the choice of an ideal person ... reflects a conception of the moral enterprise and the ideal of impartiality on which it is founded.”²⁹ Further, because her utility function satisfies the axioms of completeness and transitivity, her preferences are perfectly internally coherent; because she is fully rational and fully informed, one can only assume that her beliefs will share this degree of consistency (though coherent in terms of non-contradiction, not in possessing a rational ordering).

Gauthier’s conception of idealization is notably weaker than Rawls’s; individuals are “aware of [their] individuality” though not its particular content. Rawls’s parties only know that they represent some individuals but they do not know the identities of the individuals they might represent. For Gauthier, a rational bargaining agent is only ignorant of her ability to “identify herself as a particular person within society.”³⁰ Gauthier believes that a viable conception of idealization must be weaker than Rawls’s in order to show how each person can “identify with the ideal actor”.³¹ He thus recognizes that the ideal actor can only generate reasons for action for real-world agents if real-world agents can identify with it. The ideal actor cannot be wholly separated from the contingencies of nature: “human individuality cannot be separated from these contingencies, and moral principles must not deny but reflect them.”³² The ideal actor thus sympathizes with the agents with whom she might be identical, sharing their passions in many respects.³³ In Gauthier’s view,

²⁶ Ibid., p. 234.

²⁷ Ibid.

²⁸ Ibid.

²⁹ Ibid., p. 235.

³⁰ Ibid., p. 236.

³¹ Ibid.

³² Ibid., p. 237.

³³ It is crucial to emphasize here that Gauthier, like Rawls, is *not* identifying the reasoning and preferences of the ideal actor with that of any real-world actor. Instead, “[the ideal actor] cannot express a single set of preferences, to be

Rawlsian abstraction violates the separateness of persons, for “impartiality is not achieved by treating persons as if they lacked particular capacities and concerns; indeed impartiality among real individuals is violated by such treatment.”³⁴ Moreover, “Rawls violates the integrity of human beings as they are and as they conceive of themselves.... [he] succeeds in treating persons only as social instruments.”³⁵ For Gauthier, individuals identify with the ideal actor only because the ideal actor is concerned with their integrity and identity. Each person can thereby adopt the rules selected by the ideal actor because they can “trace lines of rationality from the actual perspectives from which we real persons choose to the ideal perspective afforded by the Archimedean point.”³⁶ It is instrumentally rational for real individuals to reason from the perspective of the ideal agent. Gauthier’s idealization only abstracts from individual capacities and preferences “because they are partial” and not because they are arbitrary.³⁷

Gauthier does not share Rawls’s desire for determinacy; in fact, he admits that “the ideal actor imposes only very weak and hardly controversial limitations on the range of acceptable social structures.”³⁸ Instead, political morality leaves open the choice of an enormous range of social institutions.³⁹ But he still has an ideal actor who selects common principles of justice for all and who does so on the basis of the standard axioms of rational choice which are often construed so as to include a commitment to radical idealization.

I.iii: Habermas

Habermas departs from Rawls in holding that “the justification of norms and commands requires that a real discourse be carried out and thus cannot occur in a strictly monological form” like the original position.⁴⁰ The problem for Rawls’s theory is that it “operationalizes the standpoint of impartiality in such a way that every individual can undertake to justify basic norms of his own.”⁴¹ Habermas believes that the only way to achieve public justification is through free argumentation

represented by a single utility function”; she is “not maximizing expected utility—her own or that of anyone else.” *Ibid.*, p. 243.

³⁴ *Ibid.*

³⁵ *Ibid.*, p. 254.

³⁶ *Ibid.*, p. 255.

³⁷ *Ibid.*, p. 257.

³⁸ *Ibid.*, p. 258.

³⁹ That said, Gauthier does think that the ideal actor must reject socialism and embrace some form of capitalism, even if markets are subject to substantial redistribution of wealth. *Ibid.*, pp. 261-2. This is still much less determinacy than what Rawls requires, however.

⁴⁰ Habermas 1999, p. 68.

⁴¹ *Ibid.*, p. 66.

that convinces citizens to recognize the claims others make on them. For John to make a legitimate claim on Reba, the claim must be submitted to Reba's critique. Habermas thereby interprets the idea of respect for free and equal persons to require that each participant be free to "respond with a yes or a no to reasons offered by way of justification and concern for the common good."⁴² Habermas sees Rawls's conception of public justification as trapping individuals into their own, first-personal perspectives. Instead, when engaged in the process of dialogue, individual self-understanding should reflect "a universally valid view of the world" where individuals are focused on what would be "in the equal interest of each individual."⁴³ For Habermas, the moral point of view is "embodied in an intersubjective practice of argumentation which enjoins those involved to an idealizing *enlargement* of their interpretive perspectives."⁴⁴ Rawls's veil of ignorance "constrains *from the beginning* the field of vision of parties in the original position."⁴⁵

While Habermas does not address Gauthier, he categorically rejects Gauthier's view that public justification is a bargaining process. Instead, it is crucial not to mix "the conditions necessary for the discursive generation of a rationally motivated consensus with the conditions necessary for negotiating a fair compromise."⁴⁶ Rather, those party to a discourse aim at "clarify[ing] a common interest."⁴⁷ That said, Habermas and Gauthier agree that Rawls's deliberative situation excessively abstracts from individual integrity. Thomas McCarthy argues Habermas's discourse situation is deliberately designed to "preserve both the integrity of individual persons and the web of interpersonal relations in which their identities are formed and maintained."⁴⁸

Yet like Rawls and Gauthier, Habermas embraces a radical conception of idealization. While he does not idealize individual preferences, he still gauges the justification of norms apart from citizens' contingencies and "empirical" features. For otherwise, norms cannot be binding:

... [I]f only empirical motives (such as inclinations, interests, and fear of sanctions) sustain the agreement [to accept the norm], it is impossible to see why a party to the contract should continue to feel bound to the norms when his original motives change.... [I]t follows that we cannot explain the validity claim of norms without rationally motivated agreement or at

⁴² McCarthy 1994, p. 45.

⁴³ Habermas 1995, p. 117.

⁴⁴ Ibid. Emphasis in original.

⁴⁵ Ibid., p. 118. Emphasis in original.

⁴⁶ Habermas 1999, p. 72.

⁴⁷ Ibid.

⁴⁸ McCarthy 1994., p. 47.

least to the conviction that consensus on a recommended norm could be brought about *with reasons*.⁴⁹

Discourse must occur in an idealized situation “removed from contexts of experience and action” and divorced from “all motives except that of the cooperative search for truth.”⁵⁰ Justified norms are those norms we would accept “in a counterfactual situation of undistorted communication and discourse ideally free of power relations.”⁵¹ Habermas embraces an ideal decision procedure, much as Rawls does, despite the fact that he defends a “more open procedure of an argumentative practice” that “does not bracket the pluralism of convictions and worldviews from the outset.”⁵² Instead, he embraces a discursive process of “will formation” where individual opinions can change and ultimately produce a rational consensus.⁵³ Habermas idealizes in this way in order to develop a conception of the reasonable where citizens take up a discursive, moral point of view aimed at a rational consensus concerning the interests of all. Specifically, he requires that no one should be barred from entering discourse and all should be able to present and criticize claims. Further, participants must be willing to take each other’s perspectives and the model should eliminate power differences between participants so that power inequalities do not affect the shape of the forthcoming consensus. Participants must avoid strategic reasoning and be sincere.⁵⁴ Habermas’s idealization also may require an infinite time to reach consensus.⁵⁵

Habermas is less clear than Rawls and Gauthier about what idealization requires. For instance, he claims that his ideal speech situation is “not formal in the sense that it abstracts from content” but is instead heavily “dependent upon contingent content being fed into it from the outside.”⁵⁶ Despite this, a participant in moral argumentation must not conceive of his duties as too embedded in “concrete habitual behavior” where questions concerning the good life “have always already been answered.”⁵⁷ Our conception of the good life is not appropriately included “before us

⁴⁹ Habermas 1975, pp. 104-5.

⁵⁰ *Ibid.*, pp. 106-7.

⁵¹ Beiner 1983, p. 28.

⁵² Habermas 1995, pp. 118-9.

⁵³ *Ibid.*, p. 130.

⁵⁴ Habermas 1999, pp. 65-6.

⁵⁵ Though, Habermas in fact has a “two track” theory of deliberation where political decision making has to eventually cut off deliberation to make a decision despite the fact that deliberation continues in the informal political sphere. See Habermas 1998, pp. 352-387 for some relevant discussion.

⁵⁶ Habermas 1999, p. 103.

⁵⁷ *Ibid.*, pp. 107-8.

as an abstract ‘ought’.”⁵⁸ That is, we cannot determine what our moral reasons are based on our conceptions of the good. Habermas separates “moral questions” and “evaluative questions,” the former concerning justice and the latter the good life. As we isolate the moral questions concerning justice, we achieve a moral point of view “dissociated from the local conventions and historical coloration of a particular form of life.”⁵⁹ Habermas excludes some contingent elements from the ideal speech situation but not others. It appears that we cannot deliberate based on our conceptions of the good but that our conversations can be started by contingent issues. Furthermore, unlike Gauthier and Rawls, Habermas provides no clear specification of which informational sets are available to participants in ideal discourse. While Habermas has a theory of *communicative* rationality, it is unclear the extent to which participants are perfect instrumental reasoners with respect to strategic interactions between citizens. Perfection may matter less for Habermas since he likely envisages such errors being corrected in discourse. And while Habermas may not be interested in agents having perfectly internally coherent beliefs, he *is* interested in generating a shared, coherent picture of the moral universe from deliberative consensus. Perhaps Habermas’s lack of commitment on these matters is due to his deeply dialogical conception of rationality, but it is again unclear.

II: Rationality, Information, Coherence

In this section, I will draw out the standard conception of idealization from the examples provided by Rawls, Gauthier and Habermas. The standard conception is usefully organized as having four-dimensions. This section outlines the three non-moral dimensions of idealization: rationality, information and coherence; it then situates Rawls, Gauthier and Habermas’s theories into these dimensions. It also introduces some of the theoretical costs and benefits of the standard conception of idealization. The fourth condition, reasonableness, includes constraints intended to represent an impartial or moral point of view; thus the first three dimensions are non-moral in that they do not directly concern taking a moral point of view. For this reason, I set reasonableness aside until Section III.

⁵⁸ Ibid.

⁵⁹ Ibid., p. 109.

II.i: Rationality

Public reason liberal accounts of rationality divide into three dichotomies: individual vs. collective rationality, interactive vs. non-interactive rationality, and full vs. bounded rationality. The standard conception of rationality within public reason liberalism is usually individual, non-interactive and full, though there are some exceptions.

The distinction between individual and collective rationality has been central to the public reason liberal tradition, though it is hard to state the distinction concisely. A theory of individual rationality is a theory of rationality for an individual, whereas a theory of collective rationality is a theory of reasoning from a collective or “we” perspective. Another way of drawing the distinction is that a conception of individual rationality requires that rational individuals maximize value for themselves, whereas collective rationality requires individuals to maximize value for a group.⁶⁰ It is important to stress, however, that maximizing value for a group is maximizing value by taking a “we” perspective, not merely maximizing value for a group from one’s perspective as an individual. Assessments are made from the perspective of the group. Rawls and Gauthier do not advance theories of collective rationality; Habermas’s position is less clear. The distinction between interactive and non-interactive reasoning depends upon, in Fred D’Agostino’s terms, whether the process of public justification involves “interaction between the parties” or not.⁶¹ Rawls’s theory is typical of non-interactive theories, as there is no interaction between parties to the original position; again, they all reason in the same way. But both Gauthier and Habermas’s theories are interactive, Gauthier through the bargaining process and Habermas through public discourse. While interactive and collective rationality are important topics within public reason liberalism, they are not central to our core problem in this dissertation, for they are not central elements involved in attributing reasons to individuals within the role of religion debate. The role of religion debate typically focuses on the reasoning of individuals given their own principles and projects that are not necessarily collective or developed interactively. Although, it is true that some deliberative democrats will worry that religious contributions to the public sphere are not susceptible to dialogical challenge and that making religious claims are somehow not in the spirit of public discussion.⁶² This concern leads

⁶⁰ Two of the major collective rationality theorists within political philosophy are Christopher McMahon and Philip Pettit. See McMahon 2001 and Pettit 1996. For an overview on the use of collective rationality within contemporary liberal political theory, see Gaus 2003, pp. 83-118.

⁶¹ D’Agostino 1996, p. 33.

⁶² While Habermas does not make this strong of a claim, he does argue that religious citizens, despite being able to advance religious claims in the public sphere must engage in a process of “self-modernization” where they, for instance,

directly to accessibility and shareability requirements or at least requirements of “translating” religious language into secular language, which we have already addressed (Ch. 6). For our purposes, we can set these two distinctions aside.

The distinction between theories of full rationality and less-than full rationality is central. We have already rejected populist idealization values, or conceptions of idealization that model citizens’ reasons as those they presently recognize, since ordinary citizens often reason poorly. To attribute reasons to them on the basis of poor reasoning would corrupt the process of public justification (Ch. 1, IV.ii). Public reason liberals must idealize; the only question is how much. A theory of full rationality idealizes agents by giving parties flawless cognitive power. Gerald Gaus holds that a fully rational agent “follows impeccable epistemic norms, ... changes his beliefs by making all the inferences from his current set of fully affirmed beliefs, and who employs the fullest possible information set.”⁶³ I have separated out the information condition, but the rest of Gaus’s description holds. The fully rational agent reasons perfectly. Rawls and Gauthier explicitly attribute full rationality to their agents; their parties have the reasoning power of gods. For Gauthier, the idea of full rationality requires not merely perfect reasoning but the more reasonable aim of modeling an agent with complete and transitive preferences. She can thereby be described as having weakly-ordered preferences. She acts perfectly rationally if she always acts to achieve what she most prefers as described by her utility function. Further, she is fully rational when she has the cognitive capacity to correctly process all information relevant to determining how to satisfy her preferences. Rawls is less specific but still attaches the reasoning of the parties to the standard axioms of decision theory.

Alternatives to full rationality remain undeveloped within public reason liberalism.⁶⁴ One reason for this is that the appeal of full rationality initially seems irresistible. If we can discern the reasons citizens would affirm were they perfectly rational, then *surely* those would be the true reasons of citizens. The public reason liberal wants to justify demands and coercion to all rational wills in her society. How could she do better than full rationality? To abandon full rationality seems tantamount to permitting at least some of the defects of populism into public justification. If individuals are less than fully rational, they will make mistakes and reasons will be attributed to citizens on the basis of

“take an epistemic stance toward the priority that secular reasons enjoy in the political arena. This can succeed only to the extent that they convincingly connect the egalitarian individualism and universalism of modern law and morality with the premises of their comprehensive doctrines.” Habermas 1996, p. 14. It is only in this way that religious claims can play a constructive role in the public sphere. For a compelling argument that Habermas faces the integrity objection in some ways as forcefully as do Rawlsians, see Yates 2007, pp. 886-889.

⁶³ Gaus 2011, p. 236.

⁶⁴ Though, see *ibid.*, pp. 233-56. Also see Fred D’Agostino’s “wave model” as an alternative account of convergence on common norms. D’Agostino 1996, pp. 129-149.

flawed reasoning. Given these considerations, it may seem bizarre for anyone to embrace a less than full conception of rationality. Another motivation for embracing full rationality is to generate agreement; many public reason liberals tacitly assume that when suitably idealized, rational and reasonable individuals will agree about many things they otherwise would not. For instance, Rawls assumes that parties to the original position are fully rational in order to induce agreement; were they less than fully rational, the parties may reach different conclusions.

But problems with full rationality abound. Nicholas Wolterstorff complains:

What does the fact that a person would not believe what he does believe if he were fully rational have to do with how he should actually be treated in the political sphere? It is said that since the law does not coerce my fully rational fictional counterpart, the rational autonomy of that fully rationally fictional counterpart of mine is respected. That is fine. But what does that have to do with how I, in my actually condition, am properly treated?⁶⁵

As we will see in Chapter 9, a standard criticism of full rationality is that fully rational agents may have substantially different beliefs and values from the real individuals they are taken to represent. Perhaps due to the interconnectedness of an agent's reasons, tens of thousands of his beliefs could be based upon a small set of core beliefs, such as beliefs that a particular religion is true. If a fully rational version of the agent would abandon those core religious beliefs, then one might reasonably be concerned that the reasons affirmed by the fully rational agent are disconnected from the reasons of his real-world counterpart. Ironically, while this was precisely Gauthier's complaint against Rawls, Gauthier himself missed that many contingences integral to individuality may well be based on faulty reasoning. For instance, suppose that John is torn between two of his life goals: being a good businessman and being a good father. He has convinced himself that he can do both well without sacrifice but in fact he is incorrect about this. As John confronts unexpected trade-offs between work and family, he will have to change and adapt. The reasons generated by these conflicts will motivate some of the most formative decisions and activities in his life (such as going to counseling to uncover the sources of his workaholism). This contingent feature of John's individuality is integral to modeling his reasons and even Gauthier's idealization threatens to erase this feature of John when modeling his reasons.⁶⁶

⁶⁵ Wolterstorff 2007, pp. 153-4.

⁶⁶ I develop this case in more detail in Ch. 10, I.i.

Other complaints aver that the idea of full rationality is incoherent or flawed. For instance, Gaus and John Pollock have argued that full rationality creates serious problems for certain forms of reasoning. Pollock, for instance, maintains that fully rational creatures may be unable to make decisions in cases where there is no best answer. We often face cases where there may be no optimal decision. In those cases an ideally rational agent may be paralyzed where a less than fully rational agent would not be.⁶⁷ Gaus claims that fully rational agents will not necessarily converge on a particular set of reasons as their reasoning is often path-dependent.⁶⁸ We will review the concerns raised by Wolterstorff, Pollock and Gaus in Chapter 9.

II.ii: Information

Many public reason liberals lump the rationality and information dimensions together, such as Gauthier and Gaus, but it is important to distinguish them. Not only is there a natural distinction between information and how it is processed, separating the dimensions will help to clarify what a theory of idealization is supposed to do. Chapter 10 argues that a satisfactory theory of idealization formulates models of cognition and information that mimic decision-making in real-time. Real agents must navigate the social world, which requires adaptive processing and gathering information. If they are to identify with their idealized counterparts, ideal agents may need to manifest similar abilities.

Public reason liberals add information to the belief-value set of ordinary citizens to avoid holding public justification captive to ignorance. Nevertheless, they often subtract information to avoid attributing partial reasons to citizens that bias public justification in favor of individual self-interest. Rawls, Gauthier and Habermas all pursue this strategy. Their models take for granted that idealized agents possess all general facts required to determine what reasons they have. Rawls is more specific than Gauthier or Habermas in arguing that the parties to the original position know the general laws of economics and psychology. And yet, Rawls bars citizens from knowing their conceptions of the good, whereas Gauthier resists such a move. For Habermas, citizens know few contingent facts about themselves or their conceptions of the good, yet they must process a constant stream of contingent, particular information about their society.⁶⁹ For Gauthier, citizens know all the details of their societies, including contingent information about individuals. They simply do not

⁶⁷ Pollock 2006, pp. 3-19.

⁶⁸ Gaus 2011, pp. 239-41.

⁶⁹ That said, they still know what they believe about a particular issue.

know who they are. In general, then, the pressure on public reason liberals is to *remove* partializing information and *add* general information. Public reason liberals disagree about what sets of information are partializing.

Informational idealization ranges between full information and a weaker standard of adequate or relevant information. The motivations for embracing and rejecting full information conceptions of idealization are similar to those for full rationality. A full information account is attractive since it will attribute reasons to citizens based on the maximal quantity of true propositions and no false ones. Full information is also preferred because it automatically includes an egalitarian element. If all agents have full information, they all have the same information. With less than full information, agents might have different informational sets. Finally, full information might be thought to guarantee determinacy; if agents have all of the same information, then they will be more likely to converge on similar conclusions.

However, a number of costs are associated with full information accounts of rationality. First, it is not clear what it means to be fully informed; for instance, does a fully informed agent know all of the names in the phone book? Public reason liberals will certainly want to resist an information standard that imputes an enormous amount of irrelevant information to agents. But if they do so, they need criteria of relevance and such criteria do not appear in the literature. Second, full information may be saddled with indeterminacy. If agents encounter information in different orders, they may come to different conclusions, as their processing will inevitably differ even if they are identical in all other respects. Consider an example. Suppose that Reba is a widely recognized celebrity whose personal life is on display to the public. Suppose further that she is presently involved in a high-profile court case. Potential jurors who are fully aware of her reputation would likely be dismissed on the grounds that they would be inclined to be prejudicial in hearing evidence. Reba's reputation will in some sense give jurors prior expectations about how to evaluate her behavior. Their future judgments will be affected by their prior conception of Reba's life and character. That said, if Reba's character and biography are revealed in the course of the trial as evidence is presented, the same jurors can arguably be said not to be prejudicial given the fact that they learned the information along with the fair and impartial presentation of evidence regarding her case. In Reba's case, the same juror could be prejudicial towards her or not depending on when the juror learned about Reba's character and biography. One might respond by suggesting that we could assume that agents process all information at once, but increasing the speed of processing does not eliminate the need to process information in a particular order. Suppose that we upgraded John the

juror's cognitive abilities so that he could form judgments about Reba's role in a particular court case as much as we can reasonably imagine. Even so, John must still process the information in steps. As a result, it is hard to see how any plausible cognitive upgrades will not eliminate the path-dependence of John the juror's judgments. Thus, indeterminacy may plague full information accounts of reasons since the fact that an agent like John has full information will not wholly determine which views he holds and which reasons he recognizes. An additional difficulty for full information accounts is that they may generate informational overload; humans are arguably *essentially* cognitively limited. Having all relevant information may impose cognitive burdens upon them that they are unable to handle. Perhaps one could stipulate that any fully informed, idealized agent will possess all relevant computational capacities but such an agent would arguably be so different from a human being that there is no reason to suppose that it could serve as a model for what citizens' true reasons are. Thus, full information accounts are threatened with similar problems as conceptions of full rationality.

II.iii: Coherence

The third non-moral dimension of idealization is coherence. Two beliefs or preferences are coherent at least when they are not mutually exclusive, such as when two beliefs do not contradict and when two preferences are asymmetric (where A and B are symmetric if and only if A is preferred to B and B is preferred to A). Both beliefs and values can be subject to coherence requirements, but the type of consistency required for beliefs and values differ. Consistency among beliefs seems familiar. Beliefs are inconsistent when they contradict each other, for instance. But they can also be inconsistent with respect to probability assessments (as we shall see below).⁷⁰ Two beliefs might entail different assessments of the probability of various events. Value consistency is much more complex. I do not want to assume that values must be consistent in the sense that they never conflict, that is, that all values can be commensurate on a single scale. Instead, I claim that to be practically rational, agents must order their values *simply for the purpose of making choices*. This is not to say that they make some substantive judgment about how to reconcile these values or that all values can be made sense of with "greater than" or "less than" comparative relations.⁷¹

⁷⁰ For more discussion of conceptions of belief coherence, see Gaus 1996, pp. 81-84.

⁷¹ For a discussion of the relationship between value, rule-following and utility functions that defends this more formal or minimalist approach to decision theory, see Gaus 2011, pp. 158-160.

Coherence requirements are seldom discussed within the public reason literature. One reason for this is likely that because the full rationality and full information conceptions of idealization are so widespread, the issue of coherence may simply be uncontroversial. A perfect reasoner with all relevant information and an infinite amount of processing time will arguably possess a perfectly coherent set of beliefs. Further, once an agent is fully rational, the axioms of decision theory simply *entail* that she has a fully coherent set of preferences. It is probably for this reason that Gauthier, Habermas and Rawls do not discuss coherence. Nonetheless, coherence is a distinct and important dimension when it comes to less than full conceptions of rationality and information. An agent can possess a high degree of rationality and information, often drawing good inferences, even if she begins the reasoning process with a high degree of incoherent beliefs and evidence she has yet to harmonize.

The standard conception of idealization idealizes agents such that they possess fully coherent beliefs and preferences. It is easy to specify what it means to have fully coherent preferences: preferences are coherent when they follow the axioms of rational choice. These axioms require, in short, that agents have a weak preference-ordering with complete and transitive preferences. Gauthier explicitly assumes that ideal agents have perfectly coherent preferences. Rawls arguably does as well since he has the parties select among options according to the axioms of rational choice. Habermas seems non-committal.

The idea of fully coherent beliefs is a more complicated. A set of beliefs is fully coherent when it possesses what Gaus calls *global* coherence, which obtains when all beliefs are at least mutually consistent.⁷² Often global coherence is thought to obtain when the fact that an agent affirms *p* can be explained in part by *p*'s relation to other beliefs. As David Brink puts it, "coherence is not simply a matter of logical consistency. The degree of a belief system's coherence is a function of the comprehensiveness of the system and of the logical, probabilistic, and explanatory relations obtaining among members of the belief system."⁷³ Coherent beliefs are not merely consistent beliefs, but help to explain why the other is held by the relationship of each belief to the system as a whole. Contrast this with *local* coherence, the coherence of much smaller sets of beliefs. The same relations that comprise global coherence, relations of consistency and explanation, apply to a smaller set of beliefs. Consequently, there is no hard and fast difference between global and local coherence, as the requirements of coherence can be applied to a set of beliefs no matter its size or extent. A developed

⁷² Gaus 1996, pp. 77-81.

⁷³ Brink 1989, p. 103.

theory of idealization could specify whether beliefs must be coherent on a number of different dimensions.⁷⁴ It must at least specify an acceptable range of points on these dimensions. However, the specification of coherence will also partly rely on the specification of the informational and rationality dimensions as well. For instance, as an agent's cognitive capacities and informational resources are increased, she will be more likely to resolve incoherence amongst members of her belief-value set. She will not only be able to more easily identify inconsistencies and points of incoherence but will find it relatively easier to resolve them.

The benefit of full coherence is that, like full rationality and full information, it better approximates a common ideal of a perfect cognizer and roots all reason attribution in this ideal. The costs of global coherence are the same as those for full rationality and full information. Making a normal agent's beliefs completely coherent might change a substantial number of them, in which case the concern about identity raised by Wolterstorff rears its head. Further, we shall see in Chapter 9 that global coherence raises some of the same concerns about conceptual coherence and determinacy raised by full rationality and full information.

In sum, despite their differences, Habermas, Rawls and Gauthier largely share the standard conception of idealization within public reason with respect to rationality, information and coherence. This is to say that they embrace a conception of idealization that includes agents with perfect or near perfect reasoning ability, an enormous amount of information (though varying in type) and a high degree of internal consistency amongst beliefs and preferences. The general benefit of the standard conception is that it approximates the ideal of good reasoning at the heart of public reason liberalism. The general costs are first, that the standard conception attributes reasons to citizens on the basis of cognition they cannot identify with, and second, that the standard conception is somehow incoherent. We may now move to the moral element in idealization.

III: Reasonableness

When public reason liberals idealize, they circumscribe the set of public reasons to a set of reasons that can be offered impartially. Rawls, Gauthier and Habermas all have detailed conceptions of impartiality, or "reasonableness" in Rawlsian terms. Rawls's conception of the reasonable in *A Theory of Justice* is modeled by the veil of ignorance. The parties' deliberations are guaranteed to

⁷⁴ Lawrence Bonjour understands coherence as a positive function of logical and probabilistic consistency and three other variables. Bonjour 1985, pp. 95, 98. Bonjour is discussing theories of epistemic justification in these passages, but the discussion generalizes.

deliver impartial rules because parties are deprived of any information or motives that could make their choices partial. Gauthier's ideal actors are impartial because they select rules from an Archimedean point. Habermas's ideal discourse situation delivers impartial norms because discourse is held under certain conditions, including rules that everyone may speak and all have equal bargaining power.

In one way, reasonableness is a redundant category because all three figures model reasonableness largely in terms of the information dimension by depriving them of partializing information leading them to select unfair principles. But reasonableness is not redundant, for it attributes motives and beliefs to agents that other dimensions do not. Note that Rawls's conception of the reasonable plays two roles in his theory of justice. First, reasonableness is modeled by the veil of ignorance, but a citizen of the well-ordered society is also reasonable because she possesses the following four dispositions:

- (i) A disposition to engage in public justification, or to offer justifications for her principles and abide for the justified principles of others.
- (ii) A disposition to recognize the "burdens of judgment". The burdens of judgment are properties of normative concepts and judgment that will lead reasonable people to disagree. These properties include difficulties assessing how evidence weighs in favor of one position or another, how to weigh distinct, relevant factors, how to resolve conceptual vagueness, how their social status affects their weighing of evidence and values, how to balance difference sides of an issue and how to resolve or accept value conflicts.
- (iii) A disposition to reject the repression of other reasonable points of view.
- (iv) A disposition to rely on methods of reasoning that others can share or access.⁷⁵

Citizens are idealized in this case when agents who possess these five dispositions model their reasons, dispositions that would not otherwise be assigned to them by the standards of full rationality, full information and global coherence.

The standard conception of reasonableness is close to the Rawlsian view. Public reason liberalism is rooted in the recognition of reasonable pluralism, so disposition (ii) is a feature of the

⁷⁵ I draw this summary of Rawls's conception of the reasonable from Gaus 1996, pp. 131-132. See Rawls 2005, pp. 49-52, 53-58, 60, 76, 119, 162-3, 229.

standard conception of idealization. Anyone interested in public justification must recognize the burdens of judgment. Disposition (i) is also a feature of the standard conception as all parties recognize that political morality requires public justification. Disposition (iii) seems to follow from (i) and (ii) as individuals will judge that morality requires public justification and see that the burdens of judgment require recognizing reasonable pluralism. Consequently, they will resist repressing others without a public justification; since they will recognize that others have a different point of view, they will resist repressing those points of view. Disposition (iv) is superfluous to the standard condition, as it is already represented by the rationality dimension. As we saw above (Ch. 5), disposition (v) is included in the standard conception of public reasons and, in my view, should not be classified as part of the standard conception of idealization. It belongs in the specification of R and not of P. I take it that the standard conception of idealization requires that agents be given dispositions (i)-(iii).

All genuine versions of public reason liberalism share the standard conception of reasonableness. Since public reason liberalism *per se* requires the recognition of reasonable pluralism and the Public Justification Principle bars coercion that lacks a public justification, dispositions (i) through (iii) seem to accompany a general commitment to public justification. Thus I take it that any deliberative scenario constructed will be restricted to agents that possess these dispositions. I will for this reason set discussion of reasonableness aside, as I do not dispute the standard conception of reasonableness.⁷⁶

IV: Conclusion: What's at Stake

The standard conception of idealization requires that reasons be attributed to citizens on the basis of the reasoning of fully rational, fully informed, reasonable agents with globally coherent beliefs and values. We shall see in Chapter 9 that the standard conception of idealization can be contrasted with less radical theories of idealization and ranked according to our first and second person metrics. I will argue that high degrees of idealization will strip away the contingent, yet core, reason-generating beliefs and values of citizens. Standard conceptions of idealization will therefore attribute alien reasons to citizens. By attributing alien reasons to citizens, standard conceptions of idealization will

⁷⁶ However, there are further, genuinely controversial, elements within conceptions of the reasonable within public reason liberalism. Specifically, these include the informational deprivations required to make agents reasonable. Some have complained, for instance, that the veil of ignorance is too “thick” in that it deprives agents of too much information. But we address these complaints when we address full information in Chapter 9.

permit coercing citizens on the basis of reasons they do not and could not recognize under normal conditions. Consequently, standard conceptions will rank below more moderate conceptions of idealization on the grounds that they can be used to justify violating the integrity of citizens. Similarly, while standard conceptions of idealization do not necessarily lead to agreement, the standard models offered by public reason liberals like Rawls, Gauthier and Habermas all motivate an implausible degree of agreement. I shall argue that, in effect, such radical idealization unjustifiably diminishes reasonable pluralism. Consequently, radical conceptions of idealization will count as inferior to more moderate conception on second-personal grounds as well. I will also argue that the standard conception of idealization is incoherent and thus should be rejected on its own terms.

Resolving the role of religion debate requires showing how public reason liberalism can reduce the tension between the first and second-personal aspects of morality. To do so, an alternative conception of public reason is needed. The first element of that alternative conception is the convergence conception of public reasons. But as we saw above (Ch. 7, VI.iii), convergence cannot resolve the tension by itself, as it can be coupled with a radical, integrity-violating conception of idealization. Chapters 9 and 10 provide reason to reject the standard conception and adopt a moderate conception of idealization. Together, convergence and moderate idealization can solve our central problem.

Chapter 9: Against the Standard Conception of Idealization

Chapter 8 outlined the standard conception of idealization. In this chapter, I shall reject it. Recall that the standard conception has four dimensions: rationality, information, coherence and reasonableness. Above (Ch. 8, III), I accepted a standard account of the reasonable within public reason. This chapter thus criticizes the standard conceptions of rationality, information and coherence employed by public reason liberals. But it offers these criticisms against the three dimensions as a whole—what I shall refer to as the “standard conception” of idealization. The chapter contains four sections. I first offer three arguments against the standard conception: (i) it is incoherent, (ii) it violates integrity, incurring first-personal costs and (iii) it unjustifiably restricts reasonable pluralism, incurring second-personal costs. These arguments comprise Sections I, II and III respectively. Like my arguments in Chapter 6, these three arguments are comparative in that they show that moderate conceptions of idealization are superior to the standard conception. Section IV concludes.

I: Arguments from Incoherence

In Chapter 6 I both dispelled the appearance of benefits of the standard conception of public reasons and showed how it incurs higher first and second-personal costs than alternative conceptions. To demonstrate, I showed that the standard conception was inferior to alternatives that lacked shareability, accessibility and symmetry requirements. In short, I argued that for most conceptions of public reasons that contain accessibility, shareability or symmetry requirements, there exists a conception of public reasons without the requirements that is superior to it on first and second-personal grounds. Thus, there are eligible conceptions of public reasons without accessibility, shareability or symmetry that are superior to conceptions that contain them. As a result, there is strong reason to reject each requirement.

This argumentative strategy must be modified if we are to reject the standard conception of idealization. Determining the first and second-personal cost associated with the standard conception is difficult because it is unclear how agents idealized in the standard way could function. In response, I begin the main argument of the chapter by turning this challenge into a criticism of the benefits of

standard conception. The primary reason to adopt the standard conception is that it approximates the ideal of rationality at the core of public reason liberalism. If adding information, cognitive power and coherence to a belief-value set removes errors in reasoning, public reason liberals should not resist these upgrades. Yet if the upgrades are so extreme as to make agent cognition opaque, the benefits of the standard conception are undermined. Radically idealized agents may face insoluble decision-making problems. Consequently, the standard conception will not form part of a viable version of public reason liberalism. Thus the first stage of the argument dispels the benefits of the standard conception by showing that it runs into problems that do not involve first and second-personal costs. The specific problem is that the standard conception of idealization models citizens' reasons in ways that render implausible the claim that they model the reasons of actual citizens. Consequently, when I criticize the standard conception of idealization for being incoherent, my claim is that attempts to construct a belief-value set for a real-world agent as represented by an agent idealized in the standard way fails. Public reason liberals must hold that an idealized agent is useful when it provides a plausible account of the reasons to which real citizens are committed. But I shall argue agents idealized in the standard way cannot accomplish this task. This is because the models employed by the standard conception are incoherent. We may now transition to my argument against the standard conception of idealization.

I.i John Pollock's Critique of the Standard Conception

To demonstrate the incoherence of the standard conception, I rely on arguments against classical decision theory advanced by John Pollock. Classical decision theory concerns fully informed and fully rational choices given an internally coherent set of preferences. Since classical decision theory bears significant analogies to agent choice under the standard conception, Pollock's criticisms of the former should apply to the latter. Pollock's criticism of classical decision theory holds that it rarely applies to the decision-making of real world agents. His arguments therefore provide evidence of a gulf between the cognition of real-world agents and their idealized counterparts. This distinction motivates the conclusion that the standard conception of idealization is incoherent because it cannot count as an idealization of a real-world agent at all.

Classical decision theory is equivalent to what Pollock calls the "optimality prescription", the view that a rational agent is one who always makes an optimal choice or chooses an action with the

highest expected utility.¹ Given that agents idealized in the standard way have perfect reasoning power, full information and coherent preferences and beliefs, they cannot help but make optimal choices. So why should unidealized agents care about the recommendations of the optimality requirement? Pollock suggests a possible reply:

Although theories of ideal agents are not directly about how real agents should solve decision problems, a plausible suggestion is that the rules of rationality for real agents should be such that, as we increase the reasoning power of a real agent, insofar as it behaves rationally its behavior will approach that of an ideal rational agent in the limit.²

Ideal agents always make what Pollock calls warranted choices, choices that are justified after an agent completes all possible relevant reasoning (Ch. 7, IV). Warranted choices look normative for real-world agents because they are made on the basis of better reasoning.

Pollock maintains that real agents cannot and should not always aim to make warranted choices. One reason for this is that reasoning for cognitively sophisticated agents like human beings is “non-terminating”.³ Because agents must act, they never complete all relevant reasoning. Further, real agents must reason “defeasibly”, meaning that the future acquisition of additional information or the performance of additional reasoning may rationally necessitate the agent’s changing his or her mind.⁴ Perfectly idealized agents do not reason defeasibly because they can complete all relevant reasoning given all relevant information. Since real-world agents are constrained by time, we can only require that they perform a “respectable amount” of reasoning before they act.⁵ Thus the recommendations of idealized agents will be rooted in reasoning and information that real-world agents are too constrained to utilize. This implies that real-world agents would often be *irrational* to make warranted choices. Pollock offers the following example to illustrate. Assume that P and Q are logically equivalent but that a real-world agent does not know this given her present reasoning. Now assume that the agent has reason to believe that the probability that P is true is 66%. Suppose that choosing P is both justified and warranted. Also assume that the agent is unable to determine the probability of Q, as she has not completed enough reasoning to even have a justified belief about

¹ Pollock 2006, p. 15.

² Ibid., p. 4.

³ Ibid.

⁴ Ibid.

⁵ Ibid., p. 5. Also see Gaus 2011, pp. 246-9.

the probability of Q. From this Pollock concludes, “[I]t would be irrational for the agent to accept a bet that Q is true at 2:1 odds. That choice would not be justified” despite the fact that the choice is warranted since Q is logically equivalent to P.⁶ Therefore the rational actions of real agents and ideal agents diverge. While many believe that the rationality of real agents or “real rationality” should approximate ideal rationality, Pollock argues that a theory of rational choice for a real agent is only correct if it describes how a real rational agent should act.

Second, the cognitive limitations of real agents require that we make choices about plans, rather than actions. We often imagine that our regular decision-making ranges over acts, but Pollock denies this. In many cases, we choose a plan without filling in the details about the acts required to implement the plan. We do this in part because we are often unaware of the conditions we will face when we implement the plan. Pollock argues:

For example, I don’t normally plan ahead about which traffic light to use on a particular leg of my route. I decide that in light of the flow of traffic around me as I drive. Decisions about the details of my plans are often best left until the last minute when I will know more about the circumstances in which I am executing the plan. To accommodate this, plans are typically somewhat schematic.⁷

Further, agents may be required to choose before they can figure out the details of the plan. Pollock gives the example of being invited to give a talk in Copenhagen in nine months. He may have to make a fast decision despite having no idea what flight he will take, for instance; he must decide to work out the details later when he has both “more time and a lighter cognitive load.”⁸ For this reason, real agents should adopt plans and fill out the details later, but ideal agents can immediately complete the relevant cognitive task. Real agents thus frequently make choices over plans and need reasons to choose one plan over another, rather than reasons to select an optimal act.

Real agents also cannot make optimal choices like ideal agents can. The optimality prescription requires choice over the set of all possible actions, but it does not specify how to determine the set of alternatives. A fully rational and fully informed agent *automatically* detects the set of alternatives and ranks them, but real agents must expend resources delineating the set. One specific obstacle to determining the set is determining what it is possible to do. The optimality

⁶ Pollock 2006, p. 5.

⁷ *Ibid.*, p. 11.

⁸ *Ibid.*

prescription presumes that we can perform actions “infallibly” or with “action omnipotence”.⁹ Idealized agents possess action omnipotence because they automatically know what they are able to do. But real agents cannot always do what they think they can do. Instead, real agents formulate choices between what they should *try* to do, not what they *can* do with perfect reliability. A critic could reply that ideally rational agents will face the same problem, but this is only so if the ideally rational agent has limited information. Even so, the ideally rational agent will know the probability of failure and integrate these probabilities into her expected utility function. Real agents cannot do this.

A final difficulty for the optimality prescription is that it presupposes that real agents can choose actions or plans in isolation from one another. For ideal agents, actions and plans can be disaggregated and ranked, but real agents cannot make such determinations. To illustrate, imagine trying to implement a simple plan; to make a rational choice, we must compare the expected utility of the plan to relevant alternatives. But what is the alternative? One might think that an agent should compare implementing the plan to not implementing it. But the real world agent rarely faces a binary option as she is always implementing a host of other plans. Pollock notes that “trying to execute previous plans may affect both the probabilities and the utilities employed in computing the expected utility of the new plan.”¹⁰ For instance, achieving one plan may diminish the expected utility of another. We must choose whether to implement plans in the context of the real-world decision maker. Thus the probabilities and utilities of plans are *conditional* upon the situation of the decision-maker. As Pollock argues, “there isn’t just one possible situation the decision maker might be in, because the other plans will normally have their results probabilistically.”¹¹ Thus an agent may have to choose to implement a plan not even knowing what plan should be implemented as an alternative. Pollock argues that “local” plans should be organized in accord with “master” plans. Master plans are not universal plans for all circumstances of life. Instead, they’re plans for achieving higher-order values. Pollock thinks these plans are general enough to compare in terms of expected utility. We do not optimize over master plans because there are too many. Real decision-makers must choose “by finding good plans” that can later be improved upon. Even one’s master plan changes over time. The rules of real rationality are rules directing the evolution of plans, “not rules for finding a mythical endpoint.”¹² Real agents must be *evolutionary planners* not action optimizers.¹³

⁹ Ibid., p. 144.

¹⁰ Ibid.

¹¹ Ibid.

¹² Ibid., p. 184.

Real-world agents differ dramatically from the idealized gods of the standard conception. Real agents would often be irrational to follow the recommendations of an ideal agent given their current knowledge.¹⁴ The ideal agent would make choices over actions rather than plans and so provide irrelevant recommendations that range over actions of too fine a grain. Ideal agents optimize but real agents often should not; ideal agents are action omnipotent, and real agents are not. Further, real agents live in a world of plans that cannot be clearly disaggregated and compared. They thus make choices based on conditional probabilities of staggering complexity. It seems that taking the standard of the ideally rational agent as normative for a real-world agent is wrong-headed. In fact, it runs the risk of incoherence since the context of choice for the ideally rational agent is substantially different from the contexts of choice for the real agent.

We can draw a useful analogy from F. A. Hayek. Hayek argued that the central problem of economics is figuring out how to cope with limited information and cognitive powers. Mainstream neoclassical economics, on Hayek's view, assumes that individuals have knowledge of each other's preferences and how to meet them.¹⁵ In other words, equilibrium modeling in neoclassical theory is rooted in the assumptions of full rationality and full information. But for Hayek, humans are essentially ignorant of important facts about others and the world. Economics should investigate the conditions of real human beings with limited cognitive capacities and recommend institutions that will allow them to achieve mutual benefits from cooperation. Neoclassical economics fails to ask the most foundational question: how do individuals of limited rationality, information and time cooperate effectively? Similarly, the optimality prescription and classical decision theory do not help real agents deal with cognitive limitations encountered in the real world. Part of being a good real-world agent is coping with these problems.

Importantly, one of the points of publicly justifying moral norms is that they help to solve these problems for real world agents. Generating citizens' reasons in accord with the standard conception of idealization misses perhaps the most crucial set of problems that morality is supposed to solve—to find moral norms that help us to deal with uncertainty, mistrust, ignorance and irrationality and how to cooperate in such a world for mutual benefit.¹⁶ Reason-attribution within public reason should not be rooted in the standard conception.

¹³ I shall bracket the complex matter of whether satisficing or settling for a less than optimal choice is in fact a method of optimization given cognitive limitations.

¹⁴ The irrationality is brought out most clearly in the relationship between fully idealized agents and real-world agents, though this concern persists at lower levels of idealization to a lesser extent. I thank Chad Van Schoelandt for this point.

¹⁵ Hayek 1945.

¹⁶ This is true unless we are dealing with ideal advisors, but I shall address this below.

I.iii: Ideal Advisor Accounts as an Alternative

A critic could reply to the arguments of II.i by advancing an ideal *advisor* account of reason attribution. The ideal advisor possesses all the features of the standard ideal agent, but instead of representing citizens, she simply *advises* them about what their reasons for action are. She generates an agent's reasons by asking what she would do were she in its circumstances. The problem can be understood in terms of having an agent optimize the set of rules or norms that non-ideal or non-optimizing agents should live under. The ideal adviser appeals to its own reasons to determine what non-ideal agents have most reason to endorse. Peter Railton has defended an ideal advisor account of reasons. Reasons are rooted in what Railton calls "objectified subjective interests", interests articulated by an agent idealized in the standard way; call this agent A+, an adviser to unidealized agent A. The objectified subjective interest is "not what *he* [A+] currently wants, but what he would want his nonidealized self A to want—or, more generally, to seek—were he to find himself in the actual condition and circumstances of A."¹⁷ One might wonder why we cannot ask an idealized A what *he* wants. In short, who cares about A+? The problem is that if we idealize A, the interests he endorses "might be quite different owing to the changes in idealization."¹⁸ Railton admits that A+ might have similar problems advising A, since A+ may not be able, "... to set aside entirely his desire *in his present circumstances* with regard to himself or to A in considering what he would want to want were he to be put *in the place of* his less-than-ideal self. ... His interests may perturb his psychology in ways that alter the phenomenon we wish to observe."¹⁹ Railton seems to recognize the types of challenges discussed in II.i but he does not think they are "sufficient ground for skepticism"; instead, we need a good "perturbation theory" to account for such effects.²⁰ But since Railton does not address the deep problems raised by Pollock, I submit that the problems raised in II.i do indeed provide sufficient ground for skepticism.

One might reply that the ideal advisor would recognize the Pollock problems and correct for them. Thus she could avoid the optimality prescription in the hopes of better taking A's perspective. She would instead optimize the set of rules for non-optimizing agents by taking account of their present limitations. The ideal agent would not merely imagine herself in the same environmental circumstances as her nonideal counterpart, though, but in the same *cognitive* circumstances as well.

¹⁷ Railton 2003, p. 11.

¹⁸ *Ibid.*, p. 37.

¹⁹ *Ibid.*

²⁰ *Ibid.*

There is a problem for this reply. If A+ makes recommendations for A based on reasoning from A's cognitive and environmental circumstances, then it seems that two things will happen. First, she may generate the same recommendations as does A himself. In this case, A+ has become a spare wheel. One reason that A should appeal to A+ is precisely because A+ will yield superior recommendations for A than A would for himself. Second, A+'s recommendations to A will differ from A. I submit that A+'s recommendations will not be relevantly dissimilar from those generated by a moderately idealized agent of the sort I will outline in Chapter 10. Even if A+ is idealized in the standard way, her recommendations will be severely affected once she takes the Pollock problems into account. If she recognizes the need to give A recommendations that would be rational for him to follow, then it seems that A+ herself will have generated the sort of recommendations to which a moderately idealized agent will be committed. If she gave recommendations characteristic of a radically idealized agent, the recommendations would raise the problems raised by Pollock, but if she "rounds down" her recommendations then she will make recommendations characteristic of a moderately idealized agent. In the latter case, we have no problem. The only concern is that the ideal adviser is making recommendations on the basis of an implicit model of a moderately idealized agent. If so, we do better to begin our analysis with a conception of moderate idealization that avoids appealing to advisers. The ideal advisor account of reasons cannot save the standard conception.

II: The First-Personal Costs of the Standard Conception

The standard conception of idealization seems incoherent. But there are further reasons to reject the standard conception. Section II explains how the standard conception imposes costs upon the integrity of citizens.

II.i: First-Personal Costs and Benefits

I argued above (Ch. 4, V.iii) that a cost is imposed upon citizens when they are denied opportunities to act in accord with their projects and principles. I believe that idealization can deny these opportunities to citizens. But my claim may seem confusing, since idealization itself does not bar action. Instead, it merely attributes reasons to citizens. Idealization does not require restraint, impose blame, interfere, etc. Instead, it specifies what reasons citizens have, which seems conceptually prior to determining whether citizens may act on their reasons. How then can idealization impose first-

personal costs if it does not obstruct action? I take idealization to impose first-personal costs to the extent that it requires *revision* of one's actual belief-value set. Recall that the point of idealizing is to treat citizens according to their present reasons and thereby extend their ability to live lives of integrity and cooperate with others. However, idealizing changes citizens' beliefs and values by definition, and so always somewhat deviates from their present concerns. Therefore, I think we should understand revisions to citizens' belief-value sets as having first-personal costs. To put it another way, the more citizens' reasons are altered by idealization, the less concerned they should be with the reasons recognized by their idealized counterparts. Idealization is supposed to attribute reasons to me that extend my ability to act on my core first-personal concerns, so to the extent that idealized agents have different concerns, their recommendations seem less relevant to the problems faced by ordinary citizens. Thus, first-personal costs are imposed by idealization as such.²¹ Let's expand this idea.

Following Gilbert Harman, I shall suppose that there is a presumption against revising one's belief-value set. Harman suggests the following related principles of belief revision:

Clutter Avoidance: One should not clutter one's mind with trivialities.

Interest Condition (on theoretical reasoning): One is to add a new proposition P to one's beliefs only if one is interested in whether P is true (and it is otherwise reasonable for one to believe P).²²

According to Harman, agents should avoid acquiring trivial beliefs and should accept a presumption against belief formation, meaning that we should not try to form beliefs save in reply to challenges to our interests. If we can advance our interests by acquiring new beliefs, we should. But we need a reason to acquire new beliefs since "one can handle only a limited amount of ... tentative acceptance since one can engage in only a limited amount of inquiry at one time."²³ Harman's conception of cognizing as a real agent is similar to Pollock's conception, as both want to understand good reasoning in terms of boundedly-rational agents solving concrete problems in real-time. The Clutter Avoidance Principle and Interest Condition are justified by an appeal to economizing on costs of cognition for just these sorts of agents. In this case, interests are curiosities motivated by reasons to meet particular goals. For instance, one can have an interest in one's

²¹ Costs are always imposed by revisions but the costs can be outweighed by benefits that I will discuss below.

²² Harman 1986, p. 55.

²³ *Ibid.*, p. 53.

environment from wanting to familiarize one's self with it. More relevant for our purposes, Harman postulates an "Interest in Facilitating Practical Reasoning" which holds that "If one desires E and believes M's being true would facilitate or hinder E, one has a reason to be interested in whether M is true."²⁴ We can understand this interest as an interest in extending the effectiveness of one's ability to act with integrity. For similar reasons, we have an interest in avoiding relevant inconsistencies. Harman's Interest Condition already requires making minimal changes to one's belief-value set by revising one's beliefs only as much as one's interests require. Due to its similarity with the Pollockian conception of real-world reasoning, Harman's principles of belief revision need little additional defense. Agents with limited cognitive and informational resources have an interest in belief revision but while minimizing the cost to changing their cognitive faculties. Thus, to return to our central topic, one challenge for idealization is to show that the first-personal costs associated with belief revision are worth paying.²⁵

The first-personal benefits of idealization are the extension of an individual's ability to act on her core projects and plans, or to live an integrated life. In Harman's terms, the benefit is meeting the interest in facilitating practical reasoning. In many cases, contradictory beliefs, ignorance and poor reasoning frustrate citizens' ability to achieve their ends and the benefit of revising the relevant beliefs and values is the extension of integrity. Thus, moving from populism to more moderate idealization values has significant first-personal benefits. On the other hand, the benefits of idealization will eventually diminish and then become costs. Imagine our system of beliefs and values linked in a Quinean web, with core beliefs and values at the center, and other beliefs and values located at the periphery. I take idealization to begin revision at the periphery, restructuring beliefs and values that are of less central importance to an individual's life and those that contradict beliefs and values of greater weight. But as the harmonization of one's belief-value set increases, we proceed towards the core of the web and may begin to alter and disrupt core beliefs.²⁶ Perhaps at

²⁴ Ibid., p. 55.

²⁵ One might reply that we have a regulative interest in treating one another as free and equal that justifies adding beliefs on its own. These extra beliefs won't be "clutter". Since one of the points of idealization is to achieve this regulative interest, one might worry that clutter avoidance is not an appropriate principle. Instead, perhaps there should be a presumption in favor of acquiring beliefs that help treat others as free and equal. In my view, this isn't quite right. We indeed have such an interest, but the interest at best shows that we have reason to acquire new beliefs by meeting the presumption in favor of Clutter Avoidance, not that no such presumption exists. Further, we want a principle of Clutter Avoidance to help select among conceptions of idealization that we think all have some prima facie claim to treating others as free and equal, so the regulative interest of treating others as free and equal will not help to select a single conception of idealization on its own. Thus the principle of Clutter Avoidance can help the selection process.

²⁶ I take it that "core beliefs" are correlated with "core reasons" and projects and plans and thus to the conceptions of integrity and individual ideals developed in Chapters 2 and 4. Core first-person reasons will be rooted in reasons that in

high degrees of idealization, some individuals will realize that their religion is based on faulty evidence, all things considered. A conception of idealization radical enough to undermine the core projects associated with one's religious faith would thus attribute reasons to citizens that would directly contradict their real-world core values and principles. These deep revisions raise the objections of figures like Wolterstorff and Gauthier (Ch.8, II.i). Such revisions are so significant and radical that it is unclear whether the real-world individual has any reason to identify with her idealized counterpart. After all, the counterpart could have projects and plans diametrically opposed to her own.

Perhaps core projects, principles and values are retained at high levels of idealization. But I shall argue in Section IV that increasing idealization often leads to normalization, that is, it can lead to wider agreement among agents. In fact, this is one of the motivations for radical idealization in the first place. If so, radical idealization threatens to destroy reasonable pluralism by annihilating the diverse values and aims of a free and equal citizenry. By idealizing in the standard way, citizens' projects and plans will be homogenized with the projects and plans of others. The benefits of extension become costs. The revision costs of idealization steadily increase as we idealize, whereas the benefits increase to a certain margin and then start to decrease. At some point the costs and benefits will intersect, but they will arguably intersect at a degree of idealization far below the standard conception. We can now model the first-personal costs and benefits graphically.

II.ii: Optimizing First-Personal Benefits

As we have seen (III.i), the first-personal costs of idealization are the costs of revising one's belief-value set. We need a positive reason to change our beliefs; the default position is to retain our existing commitments. The first-personal benefits of idealization are that they extend our ability to act on our projects and plans. Idealized agents are employed by philosophers to attribute reasons to citizens to act consistently with their projects and plans. If we act on our reasons, we will therefore have more integrity not less. Treating citizens in accord with such reasons will increase their opportunities to develop their integrity, not decrease them. However, this is true only to a point. I will now represent the first-personal costs and benefits of idealization, as displayed in Figure 1.

turn are based on certain general beliefs about what the good consists in, about what is just, about the purpose of one's life, and one's highest goals. These beliefs are examples of "core" beliefs that are tied to integrity and so on.

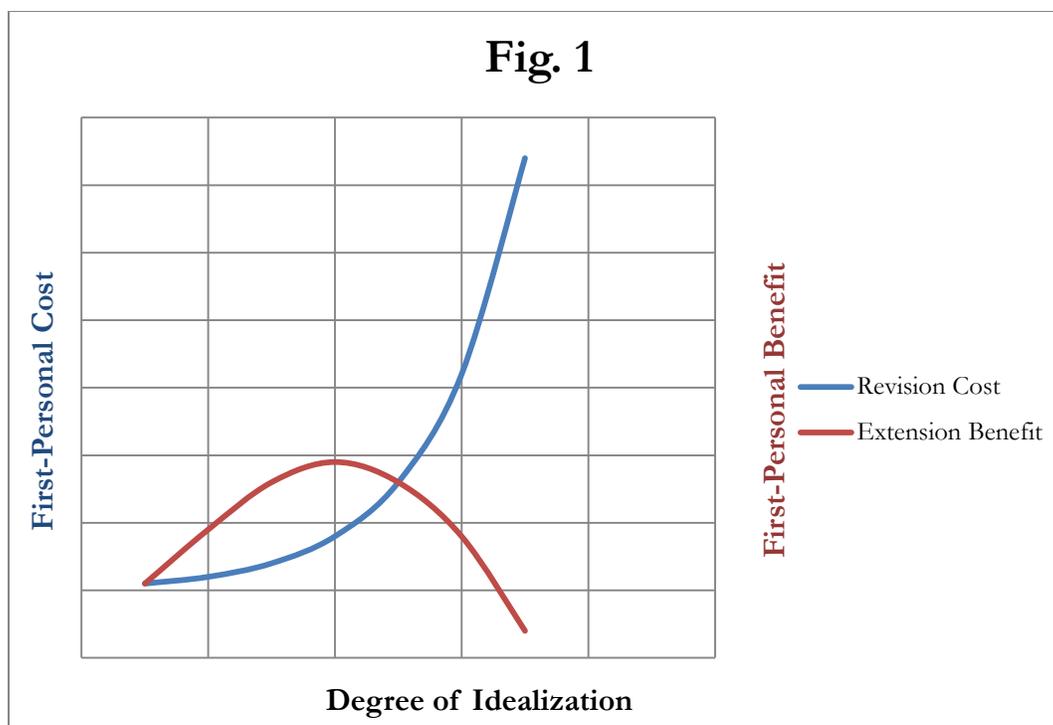


Fig. 1 displays two curves. The curve in blue is the first-personal cost curve. It represents revision costs as a function of the degree of idealization. At populist levels of idealization on the left of the curve, there are minimal or no revision costs. At moderate levels of idealization, there is a slow uptick of revision costs as peripheral beliefs and values are revised. But as revision mechanisms work their way into citizens' core beliefs and values, costs skyrocket. When a core value or belief is revised, it reverberates throughout one's peripheral beliefs and values. Further, it substantially changes the identity of a citizen, displacing vast sets of reasons. The curve in red is the first-personal benefit curve. The first-personal benefits represent the benefits of extending a citizen's reasoning as a function of idealization. As idealization increases, reasons at the periphery of an agent's belief-value set are revised, making them more consistent with her core values and principles, which is an increase in benefit. However, beyond moderate levels of idealization, core beliefs and values are revised. Consequently, the benefit of increasing a citizens' ability to act on her real-world projects and plans is sharply diminished as she is idealized so as to *lack* them.

I assume that the marginal benefits of idealization first increase and then decrease, whereas the marginal costs of revision consistently increase. As a result, the cost and benefit curves will intersect. The intersection point is the point where one additional unit of benefit achieved from idealization is no longer worth paying one additional unit of costs. At this point, the marginal

benefits of idealization equal the marginal costs. The intersection point arguably lies substantially to the left of the standard conception of idealization, as the standard conception has both revision *and* extension costs. Nonetheless, the benefits of extension push away from populism as well, moving the intersection point towards the standard conception. In this way, optimizing the benefits of idealization relative to its costs shows the superiority of a moderate conception of idealization over both populism and the standard conception. I will leave it until Chapter 10 to outline what moderate idealization involves but for now we can see how to deal with the concerns raised by Wolterstorff and Gauthier in Chapter 8. In short, when idealized in the standard way, citizens have no reason to care about the reason-attributions made by their idealized counterparts. Recall Wolterstorff's question: "What does the fact that a person would not believe what he does believe if he were fully rational have to do with how he should actually be treated in the political sphere?"²⁷ At the standard margin of idealization, the answer is "nothing". Yet at lower margins of idealization the objection *evaporates*. The reason that citizens should identify with the reason attributions derived from their moderately idealized counterparts is because their moderately idealized counterparts generate reasons that, if acted upon, will extend their real-world projects and plans and advance their real-world values more effectively.

III: The Second-Personal Costs of the Standard Conception

Public reason liberals often employ idealization to generate determinate principles of justice. Both Rawls and Gauthier use the standard conception of idealization to generate their preferred principles. Without full idealization, the parties and bargainers might disagree so much that convergence on principles of justice could not be reached. Thus idealization, public reason liberals assume, will lead to agreement or to what Fred D'Agostino called "normalization".²⁸ D'Agostino remarks that Rawls's original position reasoning is a form of normalization in the sense that members of the public find a way to choose a collectively binding mechanism of ranking proposals about how to shape the basic structure of their society. If individuals were not idealized in the standard way, they might know their identities. If they did, "it is unlikely, in the extreme, that any agreement on substantive matters would be possible."²⁹ The preferences of diverse individuals must

²⁷ Wolterstorff 2007, p. 153.

²⁸ D'Agostino 2003, p. 100. Rawls uses this term in his discussion of Rousseau in his *Lectures on the History of Political Philosophy*. Rawls 2007, p. 226.

²⁹ D'Agostino 2003

be “translated” into a collective ranking of options that makes the problem of political justification tractable.³⁰

The second-personal argument against the standard conception takes the form of a dilemma. If the standard conception normalizes too much, it will have net second-personal costs, that is, the costs of reducing reasonable pluralism will exceed the benefits of determinacy. Respecting reasonable pluralism is a bedrock commitment for public reason liberalism that can trump even the value of determinacy. On the other hand, if the standard conception fails to produce normalization, then one powerful motivation for idealizing in the standard way evaporates. Again, since determinacy is such a major benefit for a conception of public reason, a conception of idealization that reduces determinacy is less attractive, all else equal.

III.i: Second-Personal Benefits

Before proceeding to analyze the dilemma, we must ask whether idealization helps to express the values of public reason, that is, whether it has any second-personal benefits. The only second-personal benefit discussed in the literature is the putative increase in determinacy. By yielding determinate political principles, a conception of public reason enables individuals to generate a set of norms that permit all to freely advance their values; individuals can converge on norms that all can endorse. In doing so, such a conception accomplishes the goal of public justification itself—creating a society whose norms all recognize as justified. Accordingly, determinacy is a *massive* benefit because it is crucial to solving the problem of social cooperation among diverse individuals with differing ends, i.e. the central problem of social philosophy. But to justify the radical degree of normalization generated by the standard conception, we must show that determinacy can only be secured by the standard conception. The public reason literature lacks an argument to the effect that determinacy can only be achieved in this way. Further, a recent alternative conception of public reason developed by Gerald Gaus employs evolutionary mechanisms to generate determinate political principles from the reasoning of moderately idealized citizens. I cannot comment on the viability of this project but the Gaussian alternative suggests that idealizing in the standard way may not be the *only* way to achieve determinacy. In Chapter 10, I will sketch a number of ways in which conceptions of moderate idealization can achieve normalization. In this way, the second-personal benefits of complete normalization can be dispelled.

³⁰ Ibid., p. 101.

III.ii: Second-Personal Costs

If the standard conception of idealization does not produce full normalization it has no second-person benefits above and beyond those produced by less extreme conceptions of idealization that lack full normalization. It is true that *if* the standard conception of idealization is coherent it can assure us that our morality is well founded without total normalization. But the benefits acquired will not thereby exceed the net benefits of less radical conceptions of idealization. I have already suggested some reasons why the standard conception of idealization may not lead to normalization (Ch. 6, V); in short, the burdens of judgment may imply that reasonable pluralism will arise even among agents idealized in the standard way. We have also seen that the standard conception may not have unique benefits over more moderate conceptions of idealization as it may not be the only way to achieve determinacy. I will now argue that we have second-personal grounds to reject the standard conception. To establish my claim, I need only show that normalization generates problematic second-personal costs.

As we saw above (Ch. 8, I.i), Rawls's conception of idealization normalizes the parties to the original position until they agree. Such normalization raises the concern that it generates political principles by excessively restricting the considerations that are significant to citizens. There are real dangers to restricting reasonable pluralism in this way. Christopher Eberle worries that any proposed idealization threatens to attribute reasons to citizens based on the biases of the theorist whose conception of idealization is being employed. For instance, Eberle suggests that Rawlsian idealization opens the door to Rawlsians imputing their preferred political positions to citizens who would stridently oppose them. He compares appeals to improved rationality and information as appeals to "false consciousness and ideological blindness" that have been historically abused by those who cannot gain the actual assent of the coerced. Almost hyperbolically, Eberle suggests that, "one can also imagine the justificatory liberal assuring the fundamentalist parent that she'd find acceptable a policy that 'encourages' her school-age children to be more open-minded about other religions if only she were slightly better informed about what's good for her children."³¹ Idealization risks permitting the public reason liberal to "purify" the belief-value set of citizens in ways that suit her political agenda. Eberle sees no non-arbitrary way to avoid this problem,

³¹ Eberle 2002, p. 232.

Any claim about what idealized citizens would believe is uneliminably dependent on an implicit appeal to what those in much less than ideal conditions would believe, and any such appeal can—and will—be met by competing claims about what idealized citizens would believe, which are in turn dependent on an implicit appeal to a different set of claims formed in less than ideal conditions. The appeal to what idealized citizens would believe is, in short, a circuitous way for nonidealized citizens to express their disagreements with one another.³²

The danger of normalization, on Eberle's view, is not merely that reasonable pluralism is reduced but that it is reduced in a biased and oppressive fashion. When, say, Rawlsian idealize members of the public, they must make decisions about how to alter belief-value sets and these decisions risk being partial to Rawls's preferred view. As a result, the public reason liberal will simply build her values into the values of those she is idealizing; this might lead to the sort of pernicious oppression Eberle points to above. Recall that Isaiah Berlin expressed Eberle's concern long before:

For—so Hegel, Bradley, Bosanquet have often assured us—by obeying the rational man we obey ourselves: not indeed as we are, sunk in our ignorance and our passions, weak creatures afflicted by diseases that need a healer, wards who require a guardian, but as we could be if we were rational. ... In this way the rationalist argument ... has led by steps which, if not logically valid, are historically and psychologically intelligible from an ethical doctrine of individual responsibility and individual self-perfection to an authoritarian State obedient to the directives of an elite of Platonic guardians.³³

Both Eberle and Berlin see second-personal costs accompanying the standard conception of idealization. If the standard conception leads to complete normalization, it not only restricts reasonable pluralism, but turns liberal political theory into a rationale for authoritarianism. The second-personal cost of normalization is not merely that reasonable pluralism will be restricted but that such restriction will lead citizens to be coerced by the state in ways that are obviously unacceptable to them. In this way, by normalizing, public reason liberals open the door to disingenuous idealization and so permit disrespect for reasonable pluralism about the results of idealizing members of the public.

³² Ibid., p. 233.

³³ Berlin 1998, pp. 221-3.

The main motivation for idealizing in the standard way is to achieve determinacy, but the second-personal cost of radical normalization is too high. If the standard conception does not lead to full normalization—and it very well might not—then there is no additional reason (beyond first-personal reasons) to radically idealize members of the public. Thus either the standard conception imposes significant second-personal costs or it is unmotivated. That said, merely citing second-personal costs is not enough to defeat the standard conception of idealization. We must look to the relevant alternatives to show that a viable conception of idealization is in the offing. In my view, the only serious alternative to the standard conception is some form of moderate idealization. I will show in Chapter 10 that while moderate idealization gives up some determinacy, the risk of overly restricting reasonable pluralism is significantly reduced. If I am correct, given the second-personal commitments of public reason liberalism, moderate conceptions of idealization are superior to the standard conception. Coupled with the other criticisms advanced in this chapter, there will be significant reason to prefer a moderate conception of idealization to the standard conception.

IV: Conclusion

I have three aims in this chapter: (i) dispel the benefits of the standard conception by showing that it risks incoherence, (ii) demonstrate the first-personal costs of the standard conception, and (iii) demonstrate the second-personal costs of the standard conception. Section I dispelled the benefits of idealizing in the standard way by showing that it risks incoherence. Section II argued that the standard conception of idealization imposes significant first-personal costs on citizens; consequently, first-personal considerations provide strong reason to prefer moderate forms of idealization. Finally, Section III argued that idealizing in the standard way imposes significant second-personal costs on citizens; consequently, the second-personal metric also supports moderate conceptions of idealization. When applied to the problem of idealization, the dissertation's argumentative strategy provides strong reason to reject the standard conception of idealization so long as a moderate theory of idealization is viable.

Chapter 9 is the analogue of Chapter 6—it rejects the standard conception of idealization, where Chapter 6 rejected the standard conception of public reasons. In doing so, it too sets the stage for the next chapter. We saw in Chapter 7 that the convergence conception of reasons avoids the problems leveled against the standard conception of public reasons. Chapter 10 will develop a conception of moderate idealization that avoids the problems associated with the standard

conception of idealization. Thus Chapter 9 is the penultimate piece of the dissertation's puzzle. In Chapter 10, we develop the final element of the alternative conception of public reason promised in Part I. The theory of moderate idealization will complete our attempt to reconcile social morality and individual ideal in the role of religion debate and show that citizens of faith remain free within the bounds of public reason.

Chapter 10: Moderate Idealization

Part I introduced the strategy for resolving the tension between public reason liberalism and the objections advanced by its integralist critics. Part II has applied the strategy. Chapters 5-7 outlined the standard conception of public reasons, gave reason to reject it and developed the alternative convergence view. Doing so required employing first and second-personal metrics to show how the convergence conception of reasons is superior to the standard conception. Chapters 8-10 applied this strategy to theories of idealization. Chapter 8 outlined the standard conception of idealization and Chapter 9 provided reason to reject it. All that remains is to show that a viable moderate conception of idealization is superior to the standard conception.

A theory of moderate idealization must meet two criteria. First, it can admit neither gross errors in reasoning and information (the error of populism) nor can it excessively normalize citizens' reasons or attribute reasons to them incompatible with their core values, projects and principles (the error of the standard conception). Second, it should generate a determinate conception of justice in conjunction with other resources available to public reason liberalism. If moderate idealization can meet these two criteria, it is viable and will, given the arguments of Chapter 9, be superior to the standard conception of idealization. Moderate idealization can then be combined with convergence to reconcile the first and second-person points of view.

There is not enough room in the dissertation to develop a complete conception of moderate idealization, however. Instead, I will focus my efforts on delineating a range of reasonable conceptions of moderate idealization. Models of idealization within this range should all serve to resolve the conflict between the first and second-personal points of view. Defending a particular member of the set of reasonable conceptions of moderate idealization is unnecessary to demonstrate the conclusion of the dissertation as we need only show that public reason permits a range of moderate idealization models that when combined with convergence can resolve the role of religion debate.

This chapter not only develops a range of acceptable models of idealization but shows how these models satisfy the two criteria. It contains four sections. Section I draws on the Pollock-Harman conception of real-world rationality to construct a moderately idealized agent model. Section II explains how members of the range of models can converge on norms despite their diverse reasons. Section III addresses two problems for moderate idealization, and Section IV suggests several methods of discovering the reasons of citizens given the set of eligible models.

Section V concludes. By showing how moderate idealization completes the reconciliation of the first and second-personal points of view, thereby resolving the role of religion debate. In the Epilogue, we will see that in conjunction with a convergence conception of reasons, moderately idealized members of the public preserve the freedom of citizens of faith within the bounds of public reason.

I: The Moderately Idealized Agent

I.i Practical Reasoning

As we saw in Chapter 9, Pollockian agents differ from agents idealized in the standard way in several respects. They:

- (1) Rarely optimize.
- (2) Choose mid-level plans by appealing to master plans.
- (3) Engage in non-terminating reasoning.
- (4) Make defeasible inferences.
- (5) Act on justified rather than warranted choices.
- (6) Do not achieve global coherence, but instead update and harmonize beliefs in response to real world problems.

I will also assume that real-world agents have a significant degree of reasoning ability, information and coherence. It is difficult to make these standards precise, but something must be said about them. Let us begin with the ideal of rationality within moderate idealization. What it means for a moderately idealized agent to be rational is (i) to have only openly justified beliefs and (ii) to arrive at these beliefs via sound rules of inference.¹ Recall that a belief is openly justified either when it is contained within a belief-value set or when it can be derived via sound rules of inference from that belief-value set (Ch. IV, II.i). Beliefs can be inferentially and non-inferentially justified. Inferentially justified beliefs are those justified through an actual or idealized reasoning process, whereas non-inferentially justified beliefs are somehow justified immediately. An example of a non-inferentially justified belief might be one generated by a sense organ, such as a belief that it is snowing outside

¹ Note that the model of moderate idealization developed here does not require that an agent have *all* beliefs and reasons that are openly justified for it. Instead, it must have *only* openly justified beliefs and reasons.

when John looks outside of the window. Real-world agents should assign their non-inferential beliefs some degree of initial, defeasible credibility. If they do not, then it is unclear how non-inferential beliefs can be justified apart from a coherence standard, which would in turn make non-inferential beliefs inferential. Inferentially justified beliefs are those generated by justified inference rules, including the standard rules of deductive logic and some method of making inductive inferences. While we have said a bit about what good inference rules look like, we have no room to develop a theory of such rules (Ch. 9, III.i).

An informational standard is especially hard to pin down, given that different types of information will be relevant under different circumstances. For instance, moderate idealization requires a criterion of relevance, i.e., a specification of what kind of information is needed to solve a particular problem.² The dissertation has followed Gilbert Harman in understanding relevance in terms of information that will help agents meet their interests and goals (Ch. 9, III.i). If we assume that moderately idealized agents have general projects and plans, like their real-world counterparts, they will have practical interests in satisfying them. These interests, in turn, will help to specify what information is needed. A standard of information also requires a criterion of collection cost. Reasons cannot be attributed to citizens on the basis of information that they could not possibly collect. Otherwise, we might ascribe reasons to Newton by attributing knowledge of Einsteinian physics to him, knowledge that he could not be expected to construct on his own. Of course, knowing Einsteinian physics may help Newton satisfy his interests, and so meet the criterion of relevance, but a real-world agent would not regard the substantial costs required to acquire such information as worth paying. Thus, we can ascribe information to agents that requires relatively low collection costs. Processing costs also merit discussion. Sometimes information is hard to interpret and hard to reconcile with other bits of information. Thus, there may be a problem attributing reasons to citizens when a large amount of information is needed to solve a problem that, while cheaply collected, is costly to integrate. It therefore appears that moderately idealized agents can be ascribed information in accord with criteria of (i) relevance, (ii) collection cost and (iii) processing cost.

Given the standards of rationality and information developed here, we can specify a standard of coherence for beliefs and a related standard of consistency for preferences. First, we cannot attribute reasons to citizens based on a standard of global coherence of beliefs given that moderately idealized agents require time to render their beliefs consistent and only do so in response to

² This is not to say that the information required to solve a problem is delivered to the agent in a single “upgrade”. Instead, it could be delivered serially as the agent navigates the world, attempts to solve problems, fails, etc.

inconsistencies uncovered in the process of satisfying their practical interests. The set of beliefs of a moderately idealized agent will never be wholly coherent given that the agent is constantly collecting and processing new information. The agent will also act in ways that alter his or her belief-value set, as new experiences and inferences will change beliefs and values. Even if a state of global coherence could be achieved, it will be upset by the constant dynamism of a real-world agent's experiences and reasoning. Coherence can serve as a regulative ideal nonetheless. Real-world agents have an interest in having coherent beliefs and so will revise beliefs in order to make them coherent, especially when the costs of doing so are low and the interest is high.³

The proper standard of coherence is *local* coherence. Beliefs close to one another within one's network of beliefs will tend to be rendered coherent by the moderately idealized agent. Beliefs are proximate to one another when they are connected via a relatively small number of inferences. To clarify, suppose that John believes he left his front door unlocked despite the fact that his wife had just texted him that the door was locked (John often worries about whether the door is locked, so his wife is in the habit of texting him). Suppose that John presently believes that the front door is unlocked but would come to believe that the door is locked if he checks his iPhone. In this case, a moderately idealized agent that models John's reasons will believe that the door is locked since it will possess and integrate information that is relatively easy to collect, like the information encoded in John's wife's text message. Through idealization, we can harmonize John's beliefs with easily accessible information that in turn renders closely related beliefs internally consistent.

In contrast, John may have deeply entrenched beliefs that if rendered coherent would fundamentally alter his entire belief-value set. Suppose that John is both seriously committed to having a good career and to being a good father. Suppose further that John overestimates how much time and energy he will have to attend to both goals and, as a result, will pay less attention to his family than he would like. John has two aims that will often contradict one another. But it seems inappropriate to render John's beliefs coherent in this case, as eliminating the inconsistency will eliminate one of the dynamic reason-generating features of John's entire life. Due to John's conflict, he may have reason to struggle with his priorities, engage in deep personal examination and perhaps even go to therapy to combat his workaholicism or fixation on family life. Since John struggles with

³ One might wonder how beliefs can be openly justified (as demanded by our standard of rationality) and yet still contain inconsistencies. One way in which this might occur is through deriving new beliefs inferentially from different members of one's belief-value set. These inferred beliefs may then turn out to be inconsistent with one another upon further inspection. For example, engaging in extended chains of reasoning may lead an agent to develop complex solutions to mathematical problems that are in fact inconsistent, but the inconsistency could only be uncovered with an unreasonably demanding standard of reasoning for a real-world agent.

this core life conflict, he acquires many reasons to engage in these practices. Consequently, were we to harmonize the beliefs that generate John's conflict, we would eliminate many reasons that are plausibly ascribed to him. A standard of local coherence can account for this case. While John's contradictory beliefs may be local in the sense that only a small number of inferences can reveal the contradiction, John has probably maintained the contradiction over time by building up a large number of rationalizing beliefs which help convince him that he can reconcile his aims. In this way, harmonizing John's beliefs will go beyond local coherence as it will require dispensing with a vast edifice of these rationalizing beliefs. Putting the point generally, consider that many forms of mental distress are rooted in persistent, yet false beliefs that impede an agent from achieving her aims. If we idealize away these inconsistencies, we risk idealizing away the psychological features of individuals that drive their life projects.⁴

Turning to preferences, it may seem that moderately idealized agents can be ascribed a consistent set of preferences.⁵ It seems fair to assume that, in virtue of their rational faculties, moderately idealized agents will have transitive and symmetric preferences. So long as for any two exclusive choices, A and B, either A is preferred to B or B to A (symmetry), then in conjunction with transitivity, a complete preference ordering can be generated. This generalization only holds synchronically, though. As a result, moderately idealized agents may have contradictory preferences over time, particularly over long stretches of time. John, for instance, might work through his life conflict between family and business and thereby alter his preferences. Since John's preference ordering can change over time, preference contradiction is possible. These contradictions need not be problematic when evaluating particular laws in the process of being passed, since they are formulated and passed during relatively short periods of time.⁶ Any conception of idealization within political theory must assume that preferences have some kind of intertemporal consistency. Otherwise, appropriate modeling is impossible.

⁴ A concern arises here that moderately idealized agents may still have severe neuroses, that is, they will possess persistently false beliefs. If idealized agents retain these beliefs, then such beliefs can presumably be used to generate a real-world citizen's reasons. I assume that in some such cases an agent's practical interest in avoiding such neuroses will make idealizing neurotic beliefs away worthwhile. However, we might simply model proper inference rules in ways that undermine beliefs formed on the basis of neuroses. The latter is my preferred strategy, given that moderately idealized agents simply represent the structure of a citizen's belief-value set and elements related to it via inference.

⁵ I take an agent's preferences to range over states of the world to which it imputes value. These states might be comprised of a description of an agent's satiation after eating or a state where his family is happy. For our purposes the content of the states do not matter, but I suspect the content can be quite broader, so as, say, to include descriptions of procedures for arriving at fair results.

⁶ That said, dynamic preferences may pose concerns when deciding whether a constitution is publicly justified, as we may have to account for preference changes over whole lives and generations. But for now, I cannot address this complicating factor.

The moderately idealized agent engages in continuous, defeasible reasoning, possesses relevant, easily collected and processed information, has locally coherent beliefs and synchronically consistent preferences. That said, it remains unclear as to how these standards of rationality, information and coherence work in practice. The point of Section I.i has been to sketch these standards, but we can conclude on the basis of the foregoing, that moderate idealization does not wholly reconstruct citizens' reasons. The standard conception of idealization builds a representative model by constructing an agent out of a small set of motives, flawless reasoning and perfect information. In contrast, moderate idealization begins with an individual's present commitments. The constraints outlined here delimit the range of eligible idealizations. There will be many ways in which moderately idealized agents can vary within these constraints, but the constraints are sufficient to make the range of models within them plausible conceptions of idealization.

I.ii Moral Reasoning and Moral Behavior

Something must be said concerning the *moral* reasoning of the moderately idealized agent. Recall, however, that I have declined to alter the basic conception of the reasonable within the standard conception of idealization (Ch. 8, III). Instead, I will offer an account of the reasonable as I understand it. First, I will assume that moderately idealized agents act consistently with their recognized moral reasons by pursuing their individual ideals while abiding by the constraints imposed by moral reasons. Moderately idealized agents recognize that their second-personal reasons generally override their first-personal reasons. Thus, moderately idealized agents are reasonable in part because they recognize the priority of their second-personal reasons over first-personal reasons. Generally, they will obey practiced moral rules if they regard them as justified. To be sure, at some high degree of first-personal cost, moderately idealized agents may violate second-personal reasons but even here in most cases confronting such strong first-personal reasons will simply lead them to regard the moral rules in question as unjustified. What matters for our purposes is that moderately idealized agents *pursue goals with integrity* and *obey rules they see as justified*. They pursue goals by pursuing projects and plans and they have integrity because they act in accord with their values and principles when pursuing their projects and plans. Further, when they regard a moral rule as justified, they obey it for its own sake. Moderately idealized agents do not obey moral rules merely because they believe that doing so advances their projects but because they believe that obeying the rule is the right thing to do. Moral rules are always obeyed for second-personal reasons (though not necessarily for

second-personal reasons alone). Since moderately idealized agents follow justified moral rules, they frequently cooperate. Moderately idealized agents wish to achieve their aims and will recognize that cooperation dramatically improves their chances of doing so. Moderately idealized agents do not cooperate unconditionally, though. Rather, they recognize that if others impose unjustified costs on them, they are not required to accept losses. Instead, they will refuse to cooperate with those who have a disposition to not cooperate.⁷ Again, these conditions apply to all eligible conceptions of the moderately idealized agent.

II: Finding Determinacy

II.i Why We Need Determinacy

We exit Section I with a model of a moderately idealized individual as a Pollock-Harman agent that pursues its goals subject to the constraint of justified moral rules. These agents contrast sharply with Rawlsian parties to the original position as their perspectives, values and projects will differ from one another. Public reason liberalism promises to develop a framework that can show all agents that some publicly justified morality is affirmed by their first-personal concerns. Yet it can only do so if it can yield determinate principles. We can hardly have a liberal settlement if we do not know its terms. If public reason liberalism has no method of generating publicly justified moralities, there are no norms upon which citizens can coordinate. This is not to say that public reason liberalism must generate determinacy through political theory alone. Rather, a conception of moderate idealization, to be viable, only needs the ability to appeal to some kind of determinacy-generating mechanism, be it theoretical or empirical. Furthermore, it need not achieve determinacy in the same way for every social challenge. No unique solution is required; we are merely trying to show that moderately idealized agents can reliably converge on *at least one* publicly justified morality.

II.ii Why Determinacy Is a Challenge for Moderately Idealized Agents

Idealizing in the standard way is attractive precisely because it is thought to generate determinacy. Rawls argues in detail that veil of ignorance-type restrictions on the parties are required to model our

⁷ My conception of conditional cooperation is similar to David Gauthier's, though my agents have independent moral reasons for actions, a sharp and deep difference. See Gauthier 1986, pp. 157-189.

considered judgments.⁸ Gauthier provides a similar argument on behalf of his idealization.⁹ But idealizing moderately does not automatically lead to agreement among agents. Since moderate idealization is rooted in the idea of open justification, idealization begins with agents' disparate beliefs and values, and so will not necessarily generate agreement as it corrects for errors in agents' reasoning and information. Because this conception of moderate idealization is rooted in public reason liberalism, it presumes reasonable pluralism and the burdens of judgment. Consequently, extended, competent reasoning will lead to disagreement even with a substantial increase in information. Radically idealized agents may come to agree, but moderately idealized agents simply will not have reasoned enough to come to agreement about many crucial matters. Accordingly, a theory of moderate idealization deliberately eschews generating determinacy from idealization alone. It does so to preserve both a robust recognition of reasonable pluralism and to preserve integrity.

In this way moderate idealization leaves us with a certain type of indeterminacy: the reasons moderately idealized agents endorse do not themselves determine which social norms are publicly justified. This raises the two problems discussed in Chapter 7, VI.ii, that there may be *too many* norms that can be justified such that it is impossible to select among them in a rational fashion or that there may be *too few* norms that all regard as an improvement on their evaluative standards. In other words, as for social norms, we need a single set, but moderately idealized agents may over or under produce such sets of norms. We may speak for a moment in terms of inconclusively justified proposals, those that are not defeated but also cannot defeat alternatives. Public reason liberalism aims for a single set of *conclusively* justified norms, that is, norms that can defeat all competing norms. A norm, however, can defeat its competitors in conjunction with a decision procedure. A single set of laws or policies can be selected by a conclusively justified decision procedure which selects among inconclusively justified options.

Both the problem of an empty set of norms and the problem of an unmanageably large set of inconclusively justified norms reveal a challenge: moderately idealized agents can rank proposals for themselves, but it is unclear how to rank proposals for society as a whole. In other words, it is unclear how to generate a social ranking of proposals from individual rankings. Even if the process of generating a social ranking could be completed without running into decision-theoretic paradoxes like those revealed by Arrow's Impossibility Theorem, the social ranking itself may not show that one option defeats all the others. Public reason liberals must appeal to decision procedures to solve

⁸ Rawls 1971, pp. 112-122.

⁹ Gauthier 1986, pp. 235-262.

these problems. Political philosophy can determine a set of proposals that have not been defeated by moderately idealized agents, but it cannot show which is socially best, that is, which is supported by the strongest set of considerations.

But since political *theory* cannot provide such a decision procedure by itself, it can pick out decision procedures *in the world* that can do the job. Accordingly, the moderate idealization theorist is faced with the challenge of generating determinacy through methods other than idealization. These methods vary, but I will suggest that convergence upon determinate recommendations can occur through four real-world social processes: (i) deliberation, (ii) bargaining, (iii) adjudication and (iv) convention-making. This means that in accepting a moderate theory of idealization, we must also accept that discovering principles of justice and social morality require combining the resources of political theory with historical and empirical work on the structure of moral and political practices within individual societies. I will examine the complex epistemological process of unearthing a publicly justified morality in Section III. For the remainder of this section, I shall outline the four formal mechanisms by which convergence might occur.

II.iii Methods of Convergence—Deliberation

Moderately idealized agents can be said to converge whenever they affirm a set of publicly justified moral rules through a social process. There are four types of convergence processes: deliberation, bargaining, adjudication and convention. Let us begin with the process most familiar to public reason: deliberation. When faced with indeterminacy in their theoretical models, nearly all public reason liberals appeal to deliberation as a method of real-world convergence. Deliberative mechanisms are those that achieve convergence on norms through explicit, communicative and cognition-heavy activity. Deliberation usually involves explicit cognition because it requires verbalizing one's thoughts in communication and listening to what others have to say. Consequently, deliberation tends to be cognitively-expensive vis-à-vis other processes, like the formation of conventions. Deliberation aims at agreement through an exchange of reasons and arguments until all agree upon a particular proposal. The process consciously aims to resolve itself into a decision about who is right and who is wrong and thereby determine what all have reason to endorse. In this way, deliberation is a powerful mechanism for convergence because it explains how public justification is achieved. The very telos of deliberation is to arrive at decisions all can agree upon.

Deliberation is the most familiar mechanism of convergence not only because it is the most often cited, but because the ideal of public reason is usually taken to imply a commitment to engage in public deliberation. After all, what is public reason if not public *reasoning*? The problem is that there are clearly other methods by which a norm can come to be publicly justified, such as through a judicial decision that selects among two inconclusively justified proposals in a conclusively justified manner. The Public Justification Principle itself does not imply an exclusive focus on deliberation then. And in fact, there is reason to suppose that deliberation itself is parasitic on other types of mechanisms, specifically on conventions, which I will discuss below (II.vi). The tie between deliberation and public justification has already been loosened in other chapters, however (Ch.2, II.ii; Ch.3, III.iii), so we can now examine other methods of convergence.

II.iv Methods of Convergence—Bargaining

The second method of convergence is bargaining. Bargaining theory explains how agents with different interests converge on common norms. Often bargaining models are employed to settle on various rules for the division of the gains of social cooperation, such as David Gauthier's distribution rule—minimax relative concession.¹⁰ For Gauthier, self-interested parties converge on norms and distributions of goods by agreeing to a distribution rule where bargainers trade under terms that minimize the maximum concession they must make in order to convince the other bargainer to cooperate. That is to say that any alternative bargain would require that at least one party give up more than *any* party in the present bargain, assuming that minimax relative concession holds. Thus if the rule holds, while John can complain that he gave up too much, everyone else can respond that under any other arrangement one of them would have had to give up *more than John* as a proportion of their initial endowment of assets. For Gauthier, the parties—his “constrained maximizers”—will refrain from defecting on other bargainers but would instead sacrifice their own immediate short-term interests and only maximize their gains subject to the constraint of a moral norm. Significantly, Gauthier's rule does not arise from deliberation. Instead, while bargainers interact, they do not engage in cognition-heavy deliberation with the aim of consensus. However, they also do not cooperate through spontaneous order processes where neither party is aware of convergence on common norms, like conventions. Instead, bargaining models of public reason are best characterized as what Fred D'Agostino terms an “economic” conception of public justification

¹⁰ Ibid., pp. 136-141.

instead of a “political” conception.¹¹ A political conception of public justification is more similar to those affirmed by deliberative democrats, since justification rules form “from a consideration of the various issues at stake and that represent the common good.”¹² This is opposed to the economic view where the result of parties’ deliberations is “a compromise between individuals’ interests.”¹³ Gauthier’s conception of public justification is “economic” because constrained maximizers are not interested in the common good but their own good. They only constrain their maximization because they believe that by internalizing the rule of minimax relative concession that they will be better off in the long-run. The bargain or convergence arises from the negotiation of self-interested parties.

While bargaining and deliberation are distinct processes, they can be placed on a continuum that ranges between cognitively-heavy methods of convergence and cognitively-light methods of convergence. Deliberation is typically cognitively-heavy or expensive. Where a particular bargaining model falls on the spectrum depends upon whether it requires a high degree of conscious processing to reach agreement on a bargain or whether the bargaining can be achieved through relatively unconscious, cognitively inexpensive interaction. If bargaining proceeds through conscious negotiation, it will employ psychological processes that are explicit and cognitively expensive, like deliberative processes. In contrast, bargaining might proceed through non-conscious, spontaneous order processes (discussed below). I should note that the distinction between cognitively expensive methods of convergence (i.e., deliberation and explicit bargaining) and cognitively cheap methods of convergence (i.e., convention-formating and tacit bargaining) is fluid, as there may well be processes less cognitively expensive than deliberation and more cognitively expensive than evolution.

Bargaining models of convergence are cognitively diverse in the sense that different models may require different types and degrees of cognitive capacity required to reach a bargain. To demonstrate the flexibility of bargaining models with respect to the cognitive resources required to reach a bargain, we might examine a new bargaining model advanced by Ryan Muldoon. Muldoon constructs a bargaining theory which recognizes that accounts of political agreement are challenged by the “problem of diverse perspectives.”¹⁴ Due to reasonable pluralism, it is unlikely that a society can generate a common perspective within which social conflicts can be resolved. Instead of reasoning based on the same considerations or even with the aim of achieving common goals, Muldoon develops a bargaining model that avoids relying on a “neutral political perspective” and

¹¹ D’Agostino 1996, p. 32.

¹² Ibid.

¹³ Ibid.

¹⁴ Muldoon 2009, p. 122.

that requires “very little agreement between the parties involved.”¹⁵ Instead, Muldoon draws an analogy between bargaining and the formation of prices. A common economic understanding of prices views them as aggregated preferences for exchange. A market price is the price for a good or service at which the largest number of buyers and sellers will willingly exchange. To settle on a price, participants in a market need not agree on common norms or values. Both buyers and sellers regard themselves as better off with surprisingly little knowledge of their bargaining partner’s beliefs, values, interests and the like. In fact, parties to market exchange need not “even discuss [their] views on fairness: the exchange either happens or it doesn’t.”¹⁶ Bargainers can choose to converge on an exchange point or price so long as they regard themselves as better off from their own perspectives after the exchange occurs. Citizens engage one another from their own points of view. Muldoon imagines that bargaining parties can produce bargains concerning not goods and services but bundles of rights that carve up social space. Rights partition social space in that they permit citizens to act in certain ways and assign them the authority to demand that other citizens not interfere with their actions within the permitted range. Bundles of rights therefore assign individuals a sphere of private influence to make their own preferred decisions. One might find the idea of exchanging rights a bit odd. Nonetheless, to exchange rights, bargainers need only have complete preference orderings over cases of rights conflict and how they should be resolved. Bargainers can then form prices based on the trade-offs that they will make between bundles of rights. Bargainers, Muldoon argues, will tend to converge on a Nash Bargaining Solution where the outcome of the bargain is the product of the maximized gains of utility.¹⁷ Prices for bundles of rights therefore form. And the prices can be represented as social norms. These social norms will be those that maximize an individual’s space for social choice “subject to being constrained by others.”¹⁸

Muldoon’s model is unique since it permits bargainers to divide up “state space” or social space in different ways. Bargainers evaluate exchanges of bundles of rights differently not merely because they value the bundles differently but because bargainers “categorize the world differently and so ... have different conceptions of salience.”¹⁹ Bargainers disagree about how to characterize a particular bundle of rights, so they not only disagree on values but about the content *of the thing valued*. As Muldoon argues, “Disagreement does not stem only from differences in preferences. It

¹⁵ Ibid.

¹⁶ Ibid., p. 123.

¹⁷ Ibid., p. 140. Here “utility” is understood broadly to represent preferences over bundles of rights, not goods and service.

¹⁸ Ibid., p. 141.

¹⁹ Ibid., p. 142.

can also come from differences in how people interpret the world.”²⁰ These differences in interpretation may appear to make price/norm formation impossible. But Muldoon maintains that so long as agents have their own orderings over various exchanges, theorists can describe the overall bargain among agents in terms of a “union set” of all the preferred divisions of social space. Exchanges can occur despite disagreement and so prices and norms can form despite different interpretations of rights. Muldoon’s parties represent the fact that “there is no neutral representation of state space.”²¹ In other words, there is no value-neutral method of describing distinctions between individuals’ social spaces. Despite the fact that individuals can view the same sets of rights on their own, diverse and often incompatible terms, bargainers will still come to an agreement. Muldoon’s bargain takes place despite the fact that “agents do not see themselves as engaging in the same bargain – they would describe the bargain rather differently from each other.”²²

Since parties’ categorizations of the bargain differ, they have a “complete preference ordering of rights” only from their own perspective. Thus a social ranking of these state spaces or rights assignments has only a partial ordering “across the whole ... state of the world, since [the world] includes many distinctions that the agent does not recognize.”²³ In this way, a bargaining model can permit multiple perspectives to create a bargain that assigns prices to rights allocations. As Muldoon argues, “Even if different agents interpret rights differently, a price can be established, since everyone has a defined utility function.”²⁴ The price of a right is its “social cost” or “the utility loss incurred by the conflicts with other rights bundles.”²⁵ A price model,

... allows agents with different perspectives to engage with each other in a way that doesn’t privilege any given perspective. Since price is just the social cost of a right, this can be evaluated from the perspective of any given agent – if the agent is losing more than she gains, then she pulls out of the agreement. *But she never has to agree on what it is that she is bargaining about.*²⁶

²⁰ Ibid., p. 144.

²¹ Ibid.

²² Ibid.

²³ Ibid.

²⁴ Ibid., p. 145.

²⁵ Ibid., p. 146.

²⁶ Ibid., p. 147. Emphasis mine.

Muldoon's bargaining theory shows just how little information and cognitive exertion is required to converge on a common set of norms or rights. Bargainers can exchange without a common perspective or common evaluations of the benefits of norms. They do not need to know how others value those norms or their reasons for doing so. Bargainers do not even have to agree on interpretations of these norms. Compare Muldoon's bargaining model with the model of convergence reasoning developed in this dissertation. Convergence conceptions of public reasons permit a vast diversity of reasons into the public justification of political norms, so much so that the informational challenges of producing explicit, deliberative public reasoning will be impossible (Ch. 3, III.iii). There need be no strong publicity condition, where everyone knows why norms are justified to everyone else. Instead, convergence can be reached despite a wide diversity of views and interpretations of norms. Combining Muldoon's model with the model of moderate idealization, we can see how agents with *vastly* distinct sets of reasons and values can converge on common norms. Thus, bargaining models can be very helpful in generating convergence on norms even with moderately idealized agents whose values and interpretations differ. We can conclude that moderately idealized agents can converge on norms via bargaining as well as deliberation.

II.v Methods of Convergence—Adjudication

A third method of convergence is adjudication. A paradigm case of adjudication is a referee or a judge. In cases where an open dispute does not reach resolution, referees and judges are often appointed to resolve them in an authoritative way. If two self-interested sports teams dispute a ruling, they are unlikely to ever convince one another on the proper call. So they agree at the outset to submit their decisions to a referee so as to resolve their disputes. Gerald Gaus has argued that political institutions often play this role, since on his view political deliberation "is typically inconclusive" and we therefore require a method "to end the dispute, to bring it to practical closure even though we are still confronted with an open question." What is needed is a judgment made with *authority*, that is, we need a "verdict".²⁷ Gaus's "adjudicative democracy" holds that the great social contract theorists held that the "clash of private judgments characteristic of the state of nature ... [requires] an umpire who resolves disputes through the rule of law."²⁸ Unlike deliberation, adjudication cuts off discussion regardless of whether a consensus is reached. For practical purposes,

²⁷ Gaus 1996, p. 272.

²⁸ *Ibid.*, p. 275.

each views a suboptimal solution as better than no solution at all. Similarly, adjudication also ends a bargaining process in cases of dispute. The point of adjudication is to generate an authoritative decision in lieu of agreement (on bargaining or deliberation).

Adjudicative mechanisms acquire their authority by satisfying the Public Justification Principle. Under reasonable pluralism, moderately idealized agents will seldom agree on which laws or policies are justified. Discussion and bargaining can resolve some of their disputes, but not all of them. In these cases, we look for a publicly justified decision procedure that can adjudicate disputes. If a decision procedure can be publicly justified then citizens will have a method of resolving disputes even when deliberation and bargaining fail. A decision procedure or judge resolves a dispute authoritatively when all regard having a judge as an improvement on their evaluative standards over resolving all disputes for themselves.

Deliberative and bargaining models of convergence run into problems when agreement cannot be reached. In both cases, disputes aim at some final agreement and in this way can generally be expected to track public justification under many circumstances. Adjudication has the same aim, so long as judges hold themselves to the terms of their office as delineated by members of the public. Adjudication has the advantage of being able to reliably render decisions, whereas bargaining and deliberation do not have this ability. However, while adjudication has advantages, it too is subject to dispute. Moderately idealized agents may argue about “the justifiability of the umpire”.²⁹ Accordingly, some decision procedures will prove to be publicly unjustified. One might worry that while adjudication can solve problems raised by deliberation and bargaining, it cannot solve all remaining disputes because people will disagree about which decision procedure is justified. Surely we cannot agree with Hobbes that all people have reason to submit to an umpire *no matter what*, so long as the umpire (or Sovereign) does not require us to kill ourselves.³⁰ Moderately idealized members of the public will have reason to reject the authority of judges of their disputes long before that point. As Gaus puts it, “a citizen has no obligation to obey a law if the umpire is enacting defeated proposals.”³¹ At this point, adjudication must point to something deeper, a method of convergence that can hopefully resolve the problems raised by these three cognitively-heavy methods—deliberation, bargaining and adjudication.

I should point out that adjudicative mechanisms are quite diverse. We need not restrict adjudicators to a single judge or group of judges. Instead, any method of dispute resolution that

²⁹ Gaus 2003, p. 226.

³⁰ For Hobbes’ discussion of cases where the Sovereign commands a man to kill himself, see Hobbes 1994, p. 142.

³¹ Gaus 2003, p. 228.

does not rely on explicit consensus among parties (like deliberation and bargaining) but that involves a decision made by some social system can count as adjudicative. Voting rules are thus adjudicative methods. Citizens can get together and vote for a representative who will adjudicate disputes in a legislature, say, or they can even elect a judge. The point is merely that adjudicative mechanisms are numerous and diverse.

II.vi Methods of Convergence—Convention

All three methods of convergence discussed so far run into problems of continuing disagreement. While adjudication can fill in some gaps left by deliberation and bargaining, what are we to do when we disagree about decision procedures? I suggest that all three methods of convergence run into problems not merely because they cannot resolve disputes, but because they are excessively conscious and cognitively-expensive social processes. The fourth method of convergence—convention—undergirds the first three methods and can resolve the problems that they face.

Convention-formation occurs when a social practice or norm arises through a spontaneous order mechanism, a social mechanism that is, in the words of F.A. Hayek, “the result of human action but not of human design.”³² Spontaneous order processes are, in Robert Nozick’s words, those which “show how some overall pattern or design, which one would have thought had to be produced by an individual’s or group’s successful attempt to realize the pattern, instead was produced and maintained by a process that in no way had the overall pattern or design ‘in mind’”.³³ According to Nozick, there are two types of spontaneous order mechanisms by which a particular pattern can be produced: “filtering” and “equilibrium” processes. For a given pattern P, a filtering process is one that permits through “only things fitting P, because processes or structures filter out all non-P’s” whereas for equilibrium processes, “each component part responds or adjusts to ‘local’ conditions, with each adjustment changing the local environment of others close by, so that the sum of the ripples of local adjustments constitutes or realizes P.”³⁴ If we consider some social norm or practice as a case of P, then we can see how spontaneous order processes can produce convergence via filtering or equilibrium.

To give one example of a social selection mechanism, consider a famous case from the evolution of industrial standards—the competition between VHS and Betamax formats of video

³² Hayek 1973, p. 21.

³³ Nozick 1974, p. 18.

³⁴ *Ibid.*, p. 21.

tape that lasted between 1975 and 1985. Betamax died out because of the increasing returns to scale of a particular video cassette standard. Such evolution occurs with respect to moral norms as well, such as the dramatic moral shifts in attitudes about segregation and interracial marriage in the 1960s and 1970s that led to convergence on the strong anti-segregation norm prevailing in the United States today.³⁵ While the process of economic and social spontaneous order processes in both cases involved some deliberation, deliberative mechanisms were insufficient to lead to convergence on a new equilibrium. In the VHS and Betamax case, the influence of cost structures and consumer choice were the main causal contributors to convergence on the VHS format. It is hard to see how deliberation alone would have led to the choice of VHS. And while discussion almost certainly played a major role in the shift of social attitudes about race, discussion alone seems unable to explain how such convergence happened so rapidly. Racist attitudes came to be seen as morally blameworthy over the span of just a few years. Changes in affirmed moral norms often occur unconsciously or through generational changes. Further, the variations in discussion, time and audience throughout a society as large as the United States seems to suggest that something more than deliberative mechanisms were at work.

That said, the formation of conventions need not occur through spontaneous order processes. Instead, they can be achieved by deliberation, bargaining and adjudication. But what makes the formation of conventions distinct is that they can form *without appeal* to any conscious, deliberate method of convergence. This is an enormous advantage for conventions because they can form in cognitively cheap methods, without much attention by moderately idealized agents. Agents can avoid processing time and navigate the world as they otherwise would. In this way, they can end up embracing publicly justified norms through their practices. While a spontaneous order mechanism is social in that it involves interactions between members of a society, convention change can occur unconsciously and non-propositionally. No one need say, “Ah, I see that my previous reasons were poor ones. I will change my behavior.” Spontaneous change can affect an individual simply through motivating her to imitate the behavior of another. Mechanisms like imitation require far less cognition than developing an opinion about a topic, discussing it and being open to changing one’s mind.³⁶ But there is no reason for public reason liberals to restrict

³⁵ It is of course true that legal and political elements were a part of social evolution but they were surely insufficient for convergence on our present moral practices. Government coercion alone rarely results in the degree of social and moral change that accompanied the change in social attitudes about race in the United States.

³⁶ See Boyd and Richerson 2005, pp. 19-52 for a discussion, among other things, of the role of imitation in the spread of norms across populations. Also see Bicchieri 2006, p. 204.

themselves to dialogue as their sole convergence mechanism. If spontaneous order mechanisms are available to idealized agents, it would be positively irrational for them to confine themselves to cognition-heavy activity when cognitively-cheap mechanisms could achieve public justification. Recall that all else equal, moderately idealized agents economize on costs of cognition. Consequently, they will only expend cognitive costs when necessary to achieve their practical interests and will revise beliefs in ways that minimize changes in their belief-value sets. If imitation can achieve a publicly justified morality, moderately idealized agents will often prefer this method of convergence to adjudication, deliberation and bargaining.

Conventional methods of convergence are vital for two reasons. First, conventional methods make all the other methods possible. Conscious, deliberative methods of convergence can only function against a backdrop of justified conventions.³⁷ Second, conventions, by acting pre-deliberatively, can stop the regress of disagreements raised by adjudicative mechanisms. Let's explore these two advantages in turn. Using deliberation to illustrate, notice that deliberative models depend upon conventions to determine appropriate conversation topics and shared meanings which could not have been constructed through deliberation alone. As Habermas is well-aware, deliberation presupposes a background set of norms that are already agreed upon.³⁸ Following philosophers like Ludwig Wittgenstein, Habermas affirms that all disagreement occurs against a background of consensus with respect to certain subjects.³⁹ Of course, this does not mean that we cannot subject the "background" to scrutiny as well, just that for any particular discussion, certain norms are held fixed while others are in flux. Notice now that we not only share certain background norms of conversation, but for interpersonal behavior as well. For now, let us assume that, say, norms against lying, interruption, insult are justified. How could they have come to be so justified? Presumably we did not decide through dialogue alone that, say, lying is wrong. We arrived at these norms through the process of learning to engage with others, not after we already began to engage with others. Wittgenstein puts the point this way: "My life shews that I know or am certain that there is a chair over there, or a door, and so on.—I tell a friend e.g. 'Take that chair over there', 'Shut the door', etc. etc."⁴⁰ We do not construct beliefs about the world from scratch. Instead, they are simply embedded

³⁷ Chad van Schoelandt suggests that conscious, deliberative methods of convergence can function against a backdrop of *unjustified* conventions as well, but one might worry that the outcome of deliberation may maintain the imposition of unjustified norms as result. There is, of course, more to say here, but there is not space to do so.

³⁸ Habermas 1998, pp. 17-27.

³⁹ As Wittgenstein famously noted, "All testing, all confirmation and disconfirmation of a hypothesis takes place already within a system." See Wittgenstein 2003, p. 107.

⁴⁰ *Ibid.*, p. 7.

in practices that we take for granted until they are subjected to scrutiny. John Searle points out that any form of collective intentionality requires following certain norms. Without them, we could not coordinate action in order to generate collective intentional states. For instance, the use of currency requires collective intentionality—all money must be publicly recognized as money or else it actually ceases to exist as money. But to come to regard something as money, we must have already had social practices of using the money for something else.⁴¹ Some of these practices are already intentional, forming what Searle calls a “Network” of intentional states within which the collective intentionality involved in monetary practices functions. For Searle, the Network requires a “Background” or a set of non-intentional, non-representational practices and capacities.⁴² Thus even the Network requires certain social practices and capacities in order to exist at all. The point is simply that deliberative activity *qua* intentional activity requires a massive network of interrelated intentional states and non-intentional abilities and practices that give context to particular elements involved in conversation.

In a sense, what develops in conversation is simply *the tip of the iceberg* when it comes to all the elements of human mental space that might generate convergence on particular norms. For public reason liberals to confine themselves to deliberation, bargaining and adjudication is to pointlessly hamper their search for a publicly justified morality. What’s worse, public reason liberals may find it impossible to generate a common morality through explicit methods of convergence because these cognitive resources are simply too thin to construct a public morality from scratch. Cognitively explicit mechanisms require background norms and practices that are themselves non-conscious in order to form. Further, they frequently generate inconclusive resolutions to disputes. One reason public reason liberals may have missed this point is that they take public reason to apply only to politics or to the management of formal systems of adjudication, deliberation and bargaining. As a result, public reason liberals miss the vast array of resources for developing a common morality.⁴³ Any complete account of norm convergence among moderately idealized agents will therefore utilize spontaneous order convention methods of convergence.

Now that we have seen how pervasive conventions are in our social practices of deliberation, bargaining and adjudication, we can see how they solve disputes about decision procedures in

⁴¹ For an extended version of this particular argument with respect to gold currency, see Mises 1953, pp. 97-123.

⁴² For an extensive discussion of the Background, see Searle 1983, pp. 141-159. For a discussion of the relationship between the Network and the Background, see Searle 1992, pp. 175-196.

⁴³ We might group these resources under the heading of a “moral enterprise,” following Kurt Baier. See Baier 1995, pp. 245-288.

politics. Many norms of social morality develop through spontaneous order mechanisms. While politics is frequently an explicit, deliberate set of social processes, the practice of social morality need not be. One advantage of the Baier-Strawson model of social morality is that it solves the problem of N-Person public justification because conventional norms can coordinate vast numbers of people (Ch.3, III.iii). By understanding second-person reasons as reasons concerning compliance with moral rules, my account of public justification already has a place for conventional, spontaneous order forms of development. So long as a norm generated by a spontaneous order phenomenon commands the endorsement of all who are subject to it, we need not worry about how the norm originated. In this way, social norms can evolve and achieve public justification. Once they have done so, they can be appealed to by moderately idealized members of the public in order to resolve disagreements about political regimes and adjudicative mechanisms. The set of a society's publicly justified norms can extend into its *pre*-political social practices and can be utilized to resolve disputes within political practices. A regress of political disagreement can end in our practice of social morality which already contains a high degree of determinacy. While disputes about social morality are common, conventions provide resources for convergence that can substantially ameliorate worries about publicly justifying decision procedures.

What remains obscure is *how* spontaneous order mechanisms produce convergence among moderately idealized agents. The *prima facie* problem with moral convergence is that, unlike economic and biological spontaneous order processes, there is no obvious causal feedback system for these processes to operate upon. For instance, biological evolution can make use of survival and reproduction. When a trait spreads among a sub-population and leads the sub-population to grow with respect to another, a new biological trait can take over a population over time. Similarly, when prices are offered by businesses, consumer activity produces feedback through which entrepreneurs can generate prices that in turn satisfy a broader range of consumer preferences. To explain convergence among moral norms, as opposed to biological traits and market prices, we must discern some feedback mechanism.

We cannot explore all the answers, so let us focus on one such mechanism: *punishment*. Members who adopt certain moral norms can sanction, blame or coerce those who refuse to comply with a particular norm. In this way, populations that act according to certain rules can keep other members involved in certain social practices by imposing costs upon them.⁴⁴ The challenge for a

⁴⁴ See Boyd and Richerson 2005, pp. 166-188 and Bicchieri 2006, pp. 1-51 for extensive discussions of how punishment leads to norm equilibrium. Boyd and Richerson suggest that biological and cultural evolution led humans to become

punishment view is to understand how it is rational for a moderately idealized agent to expend the energy necessary to punish others when it brings the agent no obvious benefit.⁴⁵ While sanction will have the overall effect of enforcing compliance with moral rules (formed by spontaneous order processes or no), punishers will rarely see these macro-level effects and will thereby fail to recognize the benefits of punishment to the whole social order. However, our model of moral reasoning for moderately idealized agents can solve this problem because it postulates that members of the public have reason to follow moral rules independently of their self-interest. Similarly, we can also construct a model of moderately idealized agents who will engage in practices of blame and sanction in order to enforce compliance with socially recognized moral norms. As we have seen, humans are rule-followers who follow rules for second-person reasons, not merely their first-person reasons. Further, they enforce rule-compliance with sanction and coercion. In this manner, human beings expend the energy required to enforce equilibrium with respect to certain moral rules. We can assume, given how deep-set and universal punishment behavior is within human populations, that moderately idealized agents will be inclined to punish non-cooperators as well. Consequently, moderately idealized agents will not only follow rules but will often punish those who break the rules. So long as agents converge on norms they all have reason to endorse, punishment can produce publicly justified norms. Thus, public reason liberals who adopt a moderate conception of idealization can appeal to punishment, blame and sanction as feedback mechanisms for the spontaneous formation of moral norms. In some ways, the use of non-deliberative methods of convergence parallels the use of deliberative methods within public reason. Habermasians often criticize Rawlsians for excessive reliance on political theory to generate convergence.⁴⁶ With this, moderate idealization theorists can agree. However, the moderate idealization theorist can appeal to greater resources to generate convergence than the deliberative democrat.

Spontaneous order processes can generate unjustified norms as well. In this subsection, we have only explained how conventions arise, not how *justified* conventions arise. Thus we have only outlined necessary conditions for a determinate publicly justified morality. The sufficient condition is actually arriving at a public justification. The determination of whether a norm is publicly justified is

“rule-following punishers” or individuals who will in fact expend energy enforcing compliance with social norms. Those who punish do so because they somehow derive utility (in a very broad sense) from it. The reason that rule-following punisher individuals exist is because societies with rule-following punishers can avoid being invaded by defectors, that is, the population can avoid being replaced by a defector population. A society of cooperative rule-followers can out-produce and replace a society of defectors since they can protect themselves from invasion. The Boyd and Richerson model suggests that evolution has led most humans to be rule-following punishers.

⁴⁵ For a direct answer to this question, see Boyd and Richerson 2005, pp. 189-203.

⁴⁶ Habermas 1995, pp. 117-9.

complex and will be addressed in Section III. Note here, however, that whether norms are justified will not be a total mystery; that is, it will not be a matter of finding a needle of justified norms in a haystack of unjustified norms. The advantages of coordinating on a moral norm are sometimes significant merely because the norm solves the problem of social cooperation. The VHS/Betamax battle could only resolve in favor of one side because, due to increasing returns to scale, the market preferred one standard to two. So long as a norm is an improvement according to the evaluative standards of all, it can be publicly justified, and since all moderately idealized agents have an interest in achieving the benefits of social cooperation, the simple existence of a norm can become a reason to adopt it.⁴⁷ If all regard the norm as moral and it thereby produces a kind of stability, it can become a part of justified morality, as paradoxical as that might seem. In this way, what is thought to be moral can become moral overtime by leading to a “bandwagon” effect that coordinates society on a norm that is a member of an eligible set of moral norms compatible with the Public Justification Principle.⁴⁸ I admit that social pressure can sometimes give moderately idealized agents reason to comply with unjustified norms, but eligible models of moderately idealized agents already assume enough information for individuals to determine whether norms are justified to them and sufficient motivation for them to communicate their rejection of the norm to their social group. They will also be motivated to do so out of frustration and disapproval of unjustified practices. Moderately idealized agents follow social rules for second-person reasons if they regard the rules as justified. They do not follow social rules they do not regard as justified (save for pragmatic, strategic reasons). A further important point is that moderately idealized agents need not acquiesce in a sub-optimal norm if they take themselves to have reason to reject it. They can, of course, appeal to the other three convergence mechanisms in order to alter or reject the norms delivered by convention. Thus deliberation, bargaining and adjudication can be used to modify social conventions. In fact, arguably a central function of political life is to do just this, to get us out of sub-optimal or oppressive social practices.

⁴⁷ See Gaus 2011, pp. 396-406 for a discussion of the increasing returns of shared norms.

⁴⁸ See *ibid.*, pp. 395-400. Many will stridently resist this move, as it may appear to violate the is-ought distinction given that I have claimed that what “is” can *become* what ought to be. But consider an example—a norm of etiquette. We all agree to a broader moral norm of not being rude at the dinner table, but we all know that what it means to be rude at the dinner table varies from culture to culture. We also know that norms of etiquette are not all deliberately designed; instead, such norms evolve over long periods of time. And yet, we believe that we *ought* to follow such norms of etiquette. But if this is so, then we can understand the obligatory nature of a norm of etiquette as arising through cultural evolution that leads to convergence on a norm within an eligible set. Some norms will clearly be ruled out, such as burning down your host’s house before sitting down to a meal. But otherwise, the appropriate norm is determined by evolutionary forces. Perhaps moral rules can be understood on analogy.

To conclude, moderately idealized agents can converge on norms through deliberation, bargaining, adjudication and the formation of conventions. In my view, a complete theory of justice and political justification will employ moderately idealized agents combined with these four methods of convergence. We can also conclude that plausible models of convergence among moderately idealized agents cannot appeal to deliberation and cognitively expensive processes alone. Instead, moderately idealized agents should appeal to cognitively cheap methods when they will be effective. It bears emphasizing that more work must be done in order to single out an optimal conception of moderate idealization. My claim is merely that the family of reasonable liberal and moderate conceptions of idealization will appeal to all of these mechanisms.

III: Two Concerns about Moderate Idealization

III.i Instability of Idealization

Before exploring how to discern what is justified, I should address two concerns about moderate idealization. The first concern is that moderate idealization might collapse into a form of radical idealization. Christopher Eberle has argued that since any of an agent's beliefs might be "compromised by ignorance or prejudice," any belief can be subjected to idealization.⁴⁹ To deny this, Eberle claims, implies an implausible form of infallibilism about beliefs which holds that some beliefs cannot possibly be compromised. Accordingly, one who defends moderate idealization has to accept a conception of epistemic justification that permits idealizing "away from *any* of an agent's beliefs about a given policy" since all beliefs are fallible. Eberle claims that if we can idealize away from any of an agent's beliefs that we can therefore idealize away from "all of [an agent's] beliefs about a given policy" which is tantamount to radical idealization.⁵⁰ Eberle suggests the following when analyzing Gaus's conception of moderate idealization: "No matter what type of belief Gaus wishes to exempt from idealization, we will have *the same reason* [emphasis mine] to idealize away from that type of belief as we have from those about which Gaus is willing to idealize."⁵¹ However, a moderate idealization theorist can accept that any of an agent's beliefs "might be compromised" while resisting the threat of radical idealization. She need merely tease out the meaning of "might be compromised." Remember that Pollock-Harman agents only revise beliefs when it is in their

⁴⁹ Eberle 2005, p. 180.

⁵⁰ Ibid.

⁵¹ Ibid., p. 182.

practical interest to do so and they revise in ways that minimize belief revision. Pollock-Harman agents do not revise their beliefs willy-nilly but do so in response to real-world challenges; they need reason to revise and they attribute reasons to citizens on the basis of these real-world cognitive and practical demands. The fact that a belief *might* be compromised will therefore not be a reason to idealize away from it. Consequently, while any belief “might be compromised,” the specification of the relevant modality differs across beliefs. Some beliefs might be easily compromised in the sense that there is a strong interest in revising them and relatively little reason not to revise them. In contrast, we may have core beliefs so entrenched that, while they “might be compromised”, the “might” is relatively improbable. As a result, it is hard to see how moderate idealization will collapse into radical idealization since the Pollock-Harman conception of epistemic justification is not committed to idealizing from any belief (i) in the same way, (ii) at the same time or even (iii) in response to the same stimulus. Eberle’s claim that we will have “the same reason” to idealize from different types of beliefs does not apply to Pollock-Harman agents.

A critic might also worry that moderate idealization is still too vague. So far I have only demarcated a range of eligible conceptions of rationality, information and coherence between populist and radical levels of idealization. The problem is that different conceptions of moderate idealization might attribute different reasons to citizens given the plurality of ways of revising complex belief-value sets. Consequently, selecting any one conception of moderate reason-attribution seems arbitrary. If selecting a conception of moderate idealization is arbitrary, then it seems open to the forms of tampering that Eberle and Berlin find so objectionable (Ch. 9, IV.ii). Why not pick the conception that attributes reasons in the way the particular public reason liberal prefers? The concern is significant but a reply can mollify its force. Consider first that in many cases moderately idealized agents will converge on general principles of justice and methods of arbitrating disputes. Regardless of how they are moderately idealized, citizens will still have an overriding interest in securing gains from social cooperation, demanding equal treatment from their fellow citizens and resolving conflicts with their compatriots. Reason-attribution with respect to a number of crucial general matters should thus remain invariant across conceptions of moderate idealization.

III.ii Divergence among Conceptions of Moderate Idealization

Nonetheless, a severe problem looms: it seems undeniable that on a vast number of issues agents moderately idealized in different ways will disagree. Let’s delineate a range of moderate idealization

between populism and radical idealization. We might call the bounds of the range “not quite populism” and “not quite radical” or P and R. Focus on John, who is moderately idealized in two ways: John_P and John_R.⁵² John_P reasons beyond his actual state, though not very much, whereas John_R reasons for quite a long time. John_R approaches Pollockian warrant whereas John_P is closer to Pollockian justification. Now suppose that John_P and John_R disagree about whether law L is justified to John given his belief-value set. Consider whether John is rationally committed to a policy to substantially restrict greenhouse gas emissions. Unidealized John doesn’t believe in global warming. He thinks it is based on fraudulent science and so opposes the policy. John_P decides to explore the matter in-depth and reads a number of periodicals, blogs and textbooks. Eventually he achieves a level of mastery over the evidence that leads him to conclude, with most climatologists, that global warming is a genuine, partly human-caused phenomenon. As a result, he now regards the costs of the policy worth paying. Which version of John gives a better account of John’s rational commitments?

We might presume that we can treat policies whose justification depends on the selection of John_P and John_R as inconclusively justified, that is, the policies are supported by reasons but these reasons do not defeat reasons to oppose the policy (as opposed to a conclusive reason, one that defeats all opposing reasons). Let us assume that John_P takes himself to have an inconclusive reason to oppose the policy and that John_R has an inconclusive reason to support the policy. In this case, since neither John has a conclusive reason to oppose the policy in question, we can treat John himself as having both inconclusive reasons to support the policy and inconclusive reasons to oppose it. There is no big problem here. Problems arise when defeaters enter the picture. Suppose that John_P takes himself to have a conclusive reason to oppose the policy P and John_R takes himself to have a conclusive reason to support policy P. In this case, treating P as inconclusively justified seems like the wrong answer. Both Johns deny that policy P is inconclusively justified, after all. How can we justify attributing an inconclusive reason to John that *both* idealized Johns deny that he has?

Perhaps we should prefer John_P or John_R. If reasons are available to prefer either John, then cases of conflict could be easily resolved. It is hard to see what these reasons would be. The range between John_P and John_R has been delineated by Chapter 9 and Chapter 10 (Sec. 1 and Sec. 2). If we

⁵² While moderate idealization has many dimensions, I will for now assume that these dimensions can be aggregated into one dimension for the purposes of discussion.

idealize too little or too much, we will unjustifiably restrict action rooted in first and second-personal reasons. The range *between* Johns is not ruled out by first and second-personal considerations. We have insufficient reason to choose which John to follow.

Perhaps we can select a conception of idealization based on the first and second-personal desiderata by locating a conception of idealization that is best on both metrics, though not so much better that it defeats all alternatives. The idea would be to rank conceptions of moderate idealization according to how they fare on the first and second-personal criteria. But what would this conception of moderate idealization look like? What are its exact rationality, informational and coherence standards? Can we develop a model that will generate the appropriate determinacy when combined with the convergence mechanisms of deliberation, bargaining, adjudication and convention-formation? We have no space to answer these questions definitively, if they can be so answered. Instead, I will merely say, following Rawls, that there are many valid liberal conceptions of justice: “there are many liberalism and related views, and therefore many forms of public reason specified by a family of reasonable political conceptions.”⁵³ This chapter only delineates the range of reasonable liberal conceptions of idealization. My claim is that the standard conception of idealization is *not valid*. It is *unreasonable*. A genuine range of idealization values will be moderate and so public reason liberals will have to appeal to real-world social processes like those I described above to generate determinacy.

IV: Discerning What Is Justified

Even if our theory of moderate idealization is correct *and* we can locate a single conception of moderate idealization as the most reasonable or valid on the relevant metrics, the theory of moderate idealization is silent on a crucial question: how can we know which norms moderately idealized agents will converge upon? Discerning the reasons affirmed by radically idealized agents is often thought to be straightforward. The models normalize agents until the answers are obvious. But moderately idealized agents will disagree, so discernment is more challenging. To begin to answer the question, I will sketch four processes by which we can discern whether a norm is justified given a theory of moderate idealization. But first we need clarity about what to expect from these processes.

⁵³ Rawls 2005, p. 450.

IV.i Starting from Where We Are

We can conceive of the project of public justification in two ways. The first seeks public justification through constructing general principles of justice from our considered judgments *a priori*. Think of Rawls's political conception of justice as an example. It is derived by harmonizing our considered judgments. The political conception can then be applied to a number of important questions. This method of understanding public reason relies heavily on pure political theory to explain whether certain norms and practices are publicly justified. Rawls later became dissatisfied with this approach and conceded that the political conception had to be fully justified to all members of the public. The political conception cannot specify how it can be justified to each reasonable comprehensive doctrine: "it is left to each citizen ... to say how the claims of political justification are to be ordered, or weighed, against nonpolitical values. The political conception gives no guidance in such questions..."⁵⁴ The second conception of the project of public justification conceives of public justification as a method of resolving conflicting claims as they arise.⁵⁵ The early Rawls arguably held this view, as he claims that "perhaps the principle aim of ethics is the formulation of justifiable principles which may be used in cases wherein there are conflicting interests to determine which one of them should be given preference."⁵⁶ If we select the second method, we face a less insuperable epistemological problem. We do not need to know if every norm in our society is publicly justified. Instead, we merely need a way of testing whether a claim is publicly justified when complaints are lodged against it, that is, when we must resolve "cases of conflicting interests."⁵⁷

I take the second approach, though it raises concerns of its own. First, it may be charged with excessive conservatism. A society may not merely have a handful of oppressive norms; instead, it may be oppressive from top to bottom. If so, then it looks like my more gradualist approach bars often necessary revolutionary change. That said, all too often, the revolutionary implementation of political theories has caused more problems than it solves. Revolutionary changes to social morality pose great risks, including the risk that a society will lose its ability to coordinate on a morality altogether. In some cases, losing this ability could be a social disaster as it would return all citizens to a moral state of nature, depriving them of the ability to motivate others to follow basic social norms. It is precisely this sort of social chaos that arises during many social revolutions.

⁵⁴ Ibid., pp. 386-7.

⁵⁵ See James 2005 for an argument that Rawls conceived of himself as engaged in the second project.

⁵⁶ Rawls 1951, p. 186-7.

⁵⁷ Ibid. Gerald Gaus calls this view the "testing conception." See Gaus 2011, pp. 424-427.

A critic might also complain that the second approach is relativistic. Since the Public Justification Principle provides a method of showing that certain moral practices are immoral this charge is false. Still, a conception of public justification that combines a convergence conception of reasons with moderate idealization permits multiple publicly justified moralities. Kurt Baier describes his own approach similarly: “My account is not relativistic in the sense that we cannot tell objectively whether [the morality of a society or an individual] is sound or unsound. But it is relativistic in the sense that what are sound moral guidelines in one moral order may be unsound in another.”⁵⁸ Public reason liberalism provides good grounds to reject some moral practices, despite the fact that it will not always yield a single true morality. Some may be concerned that even this form of relativism is problematic, but so long as a moral theory provides a method of ruling out moralities that restrict the freedom and equality of persons I fail to see the problem. Moral norms vary across societies but share many common elements. The form of moral justification defended here provides a method of criticizing these norms without imposing an abstract uniformity on them.

IV.ii Deliberation and Dialogue

We may now turn to methods of assessment. It is important to emphasize that these methods apply to *real-world citizens* to help them discern what their moderately idealized counterparts would affirm. The most obvious method is deliberation among citizens. While I criticized approaches to public reason that employ deliberation as their primary method of public justification, it is still a useful tool for assessing complaints about seemingly publicly justified norms. Deliberation can test the validity of norms already practiced by helping citizens articulate their reasons. It can aid in the construction of replacement norms through discussing alternatives. One obvious way of achieving public justification is through extended dialogues that lead to agreement. While many conversations do not generate consensus, moderately idealized agents will sometimes agree. By encouraging correction by others, sharing information and exchanging arguments, real-world dialogue can sometimes approximate the reasoning of moderately idealized agents. As reasoning through dialogue improves, citizens can grow increasingly confident that they are determining whether a norm is justified according to a standard of moderate idealization since their reasoning will often be improved by information sharing and argument.⁵⁹

⁵⁸ Baier 1995, p. 274.

⁵⁹ That said, such methods may systematically fail. See Pincione and Tesón 2006.

IV.iii History and Social Change

Determining what is justified might also proceed by studying the evolution of norms in response to social changes, such as alterations in family structure, economic institutions and technology. Sometimes justified moral norms will adapt to social change. In other cases, new norms will become justified as they replace older norms. One might interpret changes in moral norms surrounding birth control in this way. As Western societies became industrialized, children ceased being economic assets and instead became major expenses for their parents. Further, the potential for having more children increased substantially as infant mortality rates plummeted. In the past, norms against birth control may have contributed to economic well-being by producing more human beings that added net value to society. As economic circumstances changed, moral norms arguably shifted to promote material well-being. In this case, permissible birth control use arose in response to a genuine need to control how many children couples produced.⁶⁰ The norm against birth control may have been justified in some times and places but the public justification of the norm seems to have broken down as the balance of citizens' reasons changed. Most major conservative religious groups did not resist such changes, indicating that there was a broad consensus in favor of birth control which in turn indicates that such a norm was publicly justified.

Nonetheless, it must be remarked that social change can occur through illegitimate power. Oppressive norms can be imposed by majorities without their realizing it. Examples arguably include bourgeois norms against the use of cigarettes even in the privacy of one's home and severely enforced norms of political correctness. History and social change will only have a tendency to track public justification under conditions that are already somewhat just. We must bear these concerns in mind when we revise the commitments of moderately idealized agents.

IV.iv Media

The development of norms through both deliberative and non-deliberative processes is often reflected in media culture, which sometimes represents the reasoning of citizens. To some extent, the market process of media consumption reflects what citizens believe to be reasonable and unreasonable. While many claim that the media largely controls the reasoning of citizens, the

⁶⁰ Lest this suggestion smack of Social Darwinism, I am only arguing that publicly justified *norms* may have been selected for by societal changes, emphatically *not* that certain social *groups* are or should be selected for given certain social conditions.

mechanisms postulated to explain control seem implausible, especially given that many forms of media are relatively easy to avoid.⁶¹ To be sure, media can push social and political evolution in different directions. For instance, the internet has made it easy to filter media whose orientation one rejects. This “echo chamber” phenomenon may be driving the divergence of moral and social norms in the United States. But to some extent, the range of discussion in popular media often reflects what is justified and what is not. In an even moderately functional media culture, reasoning can reach relatively high degrees of sophistication and even vigorous, uncivil and dishonest debate can manifest some of citizens’ reasons.⁶² It must be granted that many media cultures are highly sensationalistic and sometimes serve powerful governmental and non-governmental interest groups. Yet this may be one reason to think that a publicly justified polity will govern its media institutions according to moral rules that bar such behavior. And to some extent social criticism of media on these grounds is widespread. Media in the United States is widely distrusted and the decreasing costs of communication have led to a substantial increase in the diversity of public voices, especially on the internet. While the internet will likely produce divergence of opinion, it also *reflects* increasing divergence and allows the public to understand the extent to which Americans disagree with one another. As internet discussions progress, they can often reach relatively high levels of reasoning. The rapid speed with which bloggers fact-check one another in order to affect their relative prominence may well lead to more accurate policy discussions than television or newspaper outlets. For instance, in the last decade, a media institution known as the econoblogosphere has arisen; within it, economists of all political stripes engage in constant, rigorous argumentation. While some relatively prominent economists remain powerful voices (e.g., Paul Krugman), other economists who were relatively unknown prior to the econoblogosphere have risen to prominence through their reputation for even-handedness and careful discussion (e.g., George Mason University economist Tyler Cowen). The arguments that develop out of these discussions may approach the reasoning of moderately idealized agents. This is not to say that economics bloggers are always civil to one another or are convinced by each other’s arguments. Rather, one can get a sense for which arguments have force and which do not through observing the discussion. These arguments can, in my view, be more plausibly attributed to citizens than those that arise in many other source of explicit reasoning available in mass media. This is due largely to the fact that the econoblogosphere contains an unusual concentration of expertise, rapid feedback mechanisms and accessible

⁶¹ For one example of such a thesis, see Chomsky and Herman 2002.

⁶² My argument here simply parallels that of John Stuart Mill, but rather than claiming that open discourse can lead to knowledge of the truth, I claim only that it can help to uncover citizens’ reasons. See Mill 1978, p. 50.

communication styles. Yet one may reasonably worry that the blogosphere only provide evidence of the rational commitments of a moderately idealized sub-group in a larger society.

IV.v Pure Political Theory

The previous three epistemic methods somewhat circumscribe the purpose of pure political theory. What then is the positive role for political theory beyond the conception of public reason developed over the last nine chapters? In my view, pure political theory is useful in analyzing political concepts and employing them to interpret data gathered with the above methods. Pure political theory can also help to examine arguments in the public square and increase their rigor. But perhaps the primary power of pure political theory is in identifying a set of potentially justified (or eligible) proposals. For instance, political theory can develop a list of proposals that no one has conclusive reason to reject and then suggest mechanisms for selecting a proposal from the list. Suppose that we accept, as I have suggested, that pure political theory is best used to restructure our present practices in response to conflicts (III.i). Political theory can therefore help create a list of viable proposals and help us to reflect upon, criticize and restructure our practices in light of the list of proposals. Even abstract political theories like Justice as Fairness will remain useful. A convergence theorist who moderately idealizes can use theoretical edifices like Rawls's as heuristics to discern points of convergence among citizens. As I argued above (Ch. 7, II), the original position can function as a heuristic for selecting points of overlap among citizens reasons. These points of overlap may be sets of publicly justified norms. The point of political theory, then, would be to identify and select among points of overlap. The difficulty will be vindicating some conceptions of liberalism over others. Thus the role of political theory will still be robust and will have to sort through our social practices in order to construct models that identify points of convergence. So, there is still a role for the forms of abstraction involved in variants of public reason liberalism, though these models cannot claim that the norms they identify are uniquely justified. Further, they must begin by giving a characterization of our present moral practices before determining whether they are justified. Rawls arguably was moving in this direction toward the end of his career.⁶³ Rex Martin explicitly adopts this approach in his *A System of Rights*, as does Gaus in his recent *The Order of Public Reason*.⁶⁴ Both works attempt to describe our social practices in a comprehensive manner and then and only then

⁶³ Rawls notes in *Political Liberalism* that, "There are many liberalisms and related views, and therefore many forms of public reason specified by a family of reasonable political conceptions." Rawls 2005, p. 450.

⁶⁴ Martin 1993, pp. 323-339. Gaus 2011, pp. 424-447.

develop a method of what Martin calls “critical justification” or a story about how our present practices might be altered.

Admittedly, our survey of methods of epistemic discernment has been speculative and hasty. But that is an unfortunate consequence of the comprehensiveness of the dissertation. There is simply not space to develop a detailed theory of epistemic discernment. Some models are available, but at this point I must submit that more work must be done in determining how to resolve complaints in light of the rational commitments of moderately idealized agents. We are left with the conclusion that there are many liberalisms that can satisfy the first and second-personal metrics applied throughout the dissertation. Other work may help to settle on a particular variant, but we cannot do that work here.

V: Conclusion

Chapter 10 has sought to delineate a range of acceptable conceptions of moderate idealization. Remember that Chapter 7 defended a convergence conception of reasons but acknowledged that convergence could be combined with a radical conception of idealization that could generate the same first and second-personal concerns as consensus conceptions of reasons. The aim of the last three chapters has been to quiet that concern. Chapter 8 explained the standard conception of idealization and Chapter 9 argued that the standard conception is unreasonable according to the foundational first and second-personal values of public reason liberalism as outlined in Chapter 4 and the introduction to Part II. Chapter 10 completes this project by showing that there are a number of conceptions of idealization that satisfy the first and second-personal tests developed in Part I. We may now move to our final conclusion.

Conclusion

The traditional religion and politics debate is based on a fundamental confusion: that public reason liberalism is hostile to religious contributions to public life. I have argued at length that the convergence-moderate idealization conception of public reason belies this common assumption. This alternative model is friendly to religion in the public sphere *and* superior to the standard conception of public reason on its own terms. The argument of Part II vindicates the claim that public reason can preserve the integrity of citizens of faith, while treating all as free and equal. Chapters 5-7 rejected the standard conception of public reasons and defended convergence. But even convergence can impose significant integrity costs if combined with a radical conception of idealization. In response, Chapters 8-10 rejected the standard (radical) conception of idealization and defended a range of reasonable conceptions of moderate idealization. I conclude the dissertation by outlining the consequence of the success of Part II.

I: The Problem Revisited

In the introduction I described the problem of the dissertation as follows:

There are two fundamental sources of normativity, a first-personal source and a second-personal source, that conflict in a complex manner. Each side in the debate draws from one of these two sources. The public reason liberal draws on a second-personal source of normativity by emphasizing the need to reason from the standpoint of others, while the integralist draws on a first-personal source of normativity, the authority of the individual to live an integrated life. And yet a viable liberal political theory must integrate these two sources of normativity. How is this to be done?

I resolve the problem via five aims:

Aim 1: State precisely what the dispute *is* and why it is philosophically interesting.

Aim 2: Validate the concerns of integralists while protecting public reason.

Aim 3: Show that public reason liberalism must change in light of integralist criticism.

Aim 4: Provide a strategy for discerning how the change must proceed.

Aim 5: Execute the strategy by defending an alternative version of public reason liberalism.

I take it that Part I satisfied my first aim, as Chapters 1 and 2 outlined the two sides of the debate and Chapters 3 and 4 advanced a contractualist model which showed that the debate rests on a deep tension among types of normative reasons. Chapters 2 and 4 and Part II met the second aim. Chapter 2 defends the integrity objection and Chapter 4 deepens that defense. Part II uses integrity-based considerations to select an appropriate conception of public reason liberalism. In this way, I sought to take seriously the concerns of integralists. Chapters 6 and 9 vindicate the third aim of the dissertation, since it used integrity-based considerations to reject the standard conception of public reasons and idealization. The defense of the dissertation's strategy in Chapter 4 and the introduction to Part II meets the fourth aim. Part II vindicates the fifth aim, though I need to say why in detail.

II: The Success of the Strategy

Note how Chapter 10 answers the problem of the dissertation. If there are eligible conceptions of moderate idealization, then convergence is saved from the threat of the standard conception of idealization, and so the case for restraint is effectively undermined. It follows that the theory developed in Part II preserves integrity and respect for reasonable pluralism and thus the first and second-personal aspects of public reason liberalism are reconciled. The result is that citizens of faith retain their freedom within the bounds of public reason, as we see below.

II.i Convergence Saved

Recall from Chapter 7 that the integrity costs of public reason liberalism can be ameliorated with a convergence conception of public reasons. Nonetheless, convergence mixed with the standard conception of idealization can attribute reasons to citizens that permit coercing them independently of their projects and plans. But since the moderate conception of idealization defeats the standard conception, this problem is avoided. Convergence can perform its crucial task since moderate idealization attributes reasons to citizens based on their projects and principles. In fact, moderate idealization is the preferred conception of idealization *precisely because* it does so (Ch.9: III.ii, IV.ii). We can conclude, given the foregoing, that convergence is preserved from the threat of the standard conception of idealization and can therefore be used to resolve the role of religion debate.

II.ii Restraint Defeated

The case for restraint is defeated. Chapters 6 and 7 vindicated convergence over consensus, save for the challenge of standard conceptions of idealization, and Chapters 9 and 10 sealed this weakness. Since convergence undermines the case for restraint, and convergence is vindicated, the case against restraint is vindicated. If restraint is undermined, religious expression and conviction need not be privatized and the integrity costs of public reason liberalism are substantially reduced. This is not to say that the use of unshared and inaccessible reasons is permitted under all political circumstances. Their use is restricted within judgments made by the Supreme Court, for instance. Proposing publicly unjustified coercion is not always permitted either; in fact those who can causally impact the outcome of the law-making process should restrain themselves from advancing proposals they regard as unjustified (Ch.6:VII). But again, public reason liberalism does not eliminate all integrity costs, nor should it. Instead, the dissertation only shows that integrity costs can be reduced enough to meet integralist objections. Without restraint, the integrity objection has no teeth (Ch.2:II.i).

II.iii Integrity and Reasonable Pluralism Preserved

Without significant integrity costs or normalization, the conception of public reason liberalism developed in Part II preserves both integrity and respect for reasonable pluralism. The *point* of convergence and moderate idealization is to economize on first and second-personal costs by selecting an alternative interpretation of justificatory reasons and members of the public. Consensus conceptions of reasons impose integrity costs and restrict reasonable pluralism, as we saw in Chapter 6, as does idealizing in the standard way, which we saw in Chapter 9. Again, Chapters 7 and 10 vindicate the alternative convergence and moderate idealization standards.

II.iv The First-Personal and the Second-Personal United

Chapters 1 and 2 explained the tension between public reason liberals and integralists. In Chapter 4, we saw that a version of contractualism that includes first and second-personal reasons shows that the religion and politics conflict is an applied, political version of the conflict between these two fundamental types of moral reasons. While the two types of reasons will inevitably conflict with one another, different conceptions of public justification exacerbate or lessen the tension to varying

degrees. I then argued that we can select among conceptions of public justification according to how they address the tension. Liberalism's promise is to extend the reach of the first-personal by imposing a second-personally justified social morality on all citizens. Conceptions of public justification that extend our ability to act on first-personal reasons while preserving restraints required for a diverse social life are superior to conceptions of public justification that do not. In this way, Part II's version of public reason defeats the standard conception while reducing the tension between first and second-personal aspects of public morality.

II.v Rational Freedom Preserved

However, like Rousseau, I am not satisfied with a conception of public reason that merely minimizes the tension between private (first-personal) perspectives and a public (second-personal) perspective. Minimizing the tension is not enough to preserve human dignity, for accepting the tension is to accept restrictions on individual liberty, and “[t]o renounce liberty is to renounce being a man, to surrender the rights of humanity and even its duties.... Such a renunciation is incompatible with man's nature.”¹ As I argued in Chapters 2 and 4, if a conception of public reason liberalism sufficiently reduces the integrity costs it imposes on citizens, then the integrity costs no longer undermine the rational freedom of citizens of faith. While integrity costs remain, the integrity benefits of acting on one's best reasons outweigh them. I thus advocate a “net benefit” justification of the freedom of citizens of faith, for freedom is preserved when the integrity benefits of respecting the demands of public reason outweigh the integrity costs. Integrity is both preserved and extended. The alternative version of public reason liberalism so extends the reach of integrity that the second-personal restrictions on first-personal agency are arguably *self-legislated*. In other words, when citizens live within the bounds of public reason properly understood, *they are free*. We have therefore found a political theory that proscribes a “form of association that will defend and protect the person and goods of each associate with the full common force, and by means of which each, uniting with all, nevertheless obey only [themselves] and remain as free as before.”² To answer Rousseau, this is the fundamental problem to which convergence and moderate idealization provide the solution.

Rawls's famous question has animated this dissertation from the outset:

¹ Rousseau 1997, p. 5.

² *Ibid.*, pp. 49-50.

How is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable religious, philosophical, and moral doctrines?³

For Rawls, religious disagreement “introduces into people’s conceptions of their good a transcendent element not admitting of compromise.”⁴ In his view, “political liberalism starts by taking to heart the depths of that irreconcilable latent conflict” between these transcendent elements and a reasonable political conception of justice.⁵ The ideal of public reason promises reconciliation but we have seen throughout the dissertation that the standard conception of public reason liberalism fails to solve this fundamental problem. Instead, it has *exacerbated* it. The point of this dissertation is to show that this exacerbating element is *wholly unnecessary to public reason liberalism*. If we adopt a convergence conception of reasons and a moderate conception of idealization, Rawls’s question can be answered and Rousseau’s problem can be solved.

II.vi Public Reason Liberalism Changed

The public reason liberal cannot declare victory over the integralist. Meeting the integralist criticisms exacts a hefty price from mainstream public reason liberalism. Integralists force public reason liberalism to adapt. First, public reason liberals must do without restraint and focus less on developing a civic code or duty of civility for purifying inputs into the process of public reasoning. Instead, public reason liberals should focus more energy on the project of designing institutions that generate publicly justified laws. Second, public reason liberals must accept that pure political theory will not pick out one uniquely justified set of political principles. In its place, public reason liberalism is best used to delimit a range of solutions to moral conflicts within already existing societies. Solving these problems requires appealing to a wide range of epistemic methods for determining whether our social arrangements are just. Third, the most commonly appreciated empirical method of public justification—deliberation—must be seen as impoverished when left to itself. Deliberation alone cannot solve the problem of determining what is publicly justified. Public reason liberals must appeal to evolutionary theory, bargaining theory, history, sociology, and even cognitive psychology to discover which norms are justified.

³ Rawls 2005, p. 47.

⁴ *Ibid.*, p. xxvi.

⁵ *Ibid.*

III: Closing Remarks

This dissertation project began when I first wondered why so many people of faith were hostile to liberalism and why so many liberals were hostile to the reasoning of people of faith. After some thought, it seems to me that these hostilities have little to do with liberalism itself. Instead, liberalism has been caught in the crossfire between religious believers and secularists. Many secular philosophers, political theorists and popular pundits have claimed the liberal mantle as their own, and their religious opponents have been happy to let them, as they frequently wish to embrace anti-liberal conclusions. In my view, public reason liberalism can survive this acrimonious culture war and reclaim its place in providing and preserving cooperation and toleration among warring sects. But it can only do so if its adherents are willing to take public reason liberalism in new directions. Many will be wary of these changes. I am not.

References

- Alston, William P. "An Internalist Externalism." *Synthese*, vol. 74 (1988): 265-283.
- . *Perceiving God: The Epistemology of Religious Experience*. Ithaca: Cornell University Press, 1991.
- Aquinas, St. Thomas. *Summa Contra Gentiles, Book I: God*. New York: University of Notre Dame Press, [1268] 1975.
- Audi, Robert. *Religious Commitment and Secular Reason*. Cambridge: Cambridge University Press, 2000.
- . "Natural Reason, Natural Rights, and Governmental Neutrality Toward Religion." *Religion and Human Rights*, vol. 4 (2009): 157-175.
- Audi, Robert and Nicholas Wolterstorff. *Religion in the Public Square: The Place of Religious Convictions in Political Debate*. Lanham: Rowman and Littlefield, 1997.
- Baier, Kurt. "The Point of View of Morality." *Australasian Journal of Philosophy*, vol. 32 (1954): 104-135.
- . *The Moral Point of View: A Rational Basis of Ethics*. Ithaca: Cornell University Press, 1958.
- . *The Rational and the Moral Order*. Chicago: Open Court, 1995.
- Barrett, Justin L. *Why Would Anyone Believe in God?* New York: Alta Mira Press, 2004.
- Barry, Brian. *Justice As Impartiality*. New York: Oxford University Press, 1996.
- Bayle, Pierre. *Political Writings*, edited by Sally Jenkinson. New York: Cambridge University Press, 2000.
- Beiner, Ronald. *Political Judgment*. London: Methuen, 1983.
- Benn, Stanley I. *A Theory of Freedom*. New York: Cambridge University Press, 1988.
- Berlin, Isaiah. *The Proper Study of Mankind: An Anthology of Essays*, edited by Henry Hardy and Roger Hausheer. New York: Farrar, Straus and Giroux, 1998.
- Bicchieri, Cristina. *The Grammar of Society*. New York: Cambridge University Press, 2006.
- Bloom, Paul. *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human*. New York: Basic Books, 2005.
- Bohman, James and Henry Richardson. "Liberalism, Deliberative Democracy, and 'Reasons that All Can Accept'." *Journal of Political Philosophy*, vol. 17 (2010): 253-274.

Boniface VIII, Pope. March 16th, 2010. "Unam Sanctum." *Catholic Library*, 1302. Retrieved March 16th, 2010. From http://www.newadvent.org/library/docs_bo08us.htm.

Bonjour, Laurence. *The Structure of Empirical Knowledge*. Boston: Harvard University Press, 1985.

Boswell, John. *Christianity, Social Tolerance, and Homosexuality: Gay People in Western Europe from the Beginning of the Christian Era to the Fourteenth Century*. New York: University of Chicago Press, 1981.

Boyd, Robert and Peter Richerson. *The Origin and Evolution of Cultures*. New York: Oxford University Press, 2005.

Brecht, Martin. *Martin Luther: His Road to Reformation 1483-1521*. Chicago: Fortress Press, 1985.

Brink, David. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press, 1989.

Brower, Bruce. "The Limits of Public Reason." *Journal of Philosophy*, vol. 91 (1994): 5-26.

Calhoun, Cheshire. "Standing for Something." *Journal of Philosophy*, vol. 92 (1995): 235-260.

Chisholm, Roderick. *Theory of Knowledge*. Upper Saddle River: Prentice Hall College Division, 1989.

Chomsky, Noam and Edward S. Herman. *Manufacturing Consent: The Political Economy of Mass Media*. New York: Pantheon, 2002.

Cox, Damian, Marguerite La Caze and Michael Levin. August 13th, 2010. "Integrity." *Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*, 2008. Retrieved October 22nd, 2010. From <http://plato.stanford.edu/entries/integrity/>.

D'Agostino, Fred. *Free Public Reason: Making It Up as We Go*. New York: Oxford University Press, 1996.

———. *Incommensurability and Commensuration: The Common Denominator*. Burlington: Ashgate, 2003.

Darwall, Stephen. "Agent-Centered Restrictions from the Inside Out." *Philosophical Studies*, vol. 50 (1986): 291-319.

———. *The Second-Person Standpoint: Morality, Respect, and Accountability*. New York: Harvard University Press, 2006.

Eberle, Christopher. *Religious Conviction in Liberal Politics*. New York: Cambridge University Press, 2002.

———. "What Does Respect Require?" in *Religion in the Liberal Polity*, edited by Terence Cuneo. Notre Dame: University of Notre Dame Press: 173-194.

——— (2011). *Religion, Conscientious Engagement and Respect*: forthcoming.

Feinberg, Joel. *Harm To Others: The Moral Limits of the Criminal Law*. New York: Oxford University Press, 1987.

Fichte, J. G. *Foundations of Natural Right*. New York: Cambridge University Press, 2000.

Fish, Stanley (1996). Why We Can't All Just Get Along. *First Things*: 18-26.

Flaubert, Gustave. *Sentimental Education*. London: Penguin Books, 2004.

Frankfurt, Harry. "Identification and Wholeheartedness." in *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, edited by Ferdinand Schoeman. New York: Cambridge University Press: 27-45.

Freeman, Samuel. "The Burdens of Public Justification: Constructivism, Contractualism and Publicity." *Politics, Philosophy and Economics*, vol. 6 (2007): 5-43.

Gagnon, Robert. *The Bible and Homosexual Practice*. New York: Abingdon Press, 2004.

Galston, William. *Liberal Purposes: Goods, Virtues, and Diversity in the Liberal State*. New York: Cambridge University Press, 1991.

Gaus, Gerald. *The Modern Liberal Theory of Man*. New York: St. Martin's Press, 1983.

———. *Value and Justification: The Foundations of Liberal Theory*. Cambridge: Cambridge University Press, 1990.

———. *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. New York: Oxford University Press, 1996.

———. *Social Philosophy*. Armonk: M.E. Sharpe, 1999.

———. *Political Concepts and Political Theories*. Boulder: Westview, 2000.

———. *Contemporary Theories of Liberalism: Public Reason as a Post-Enlightenment Project*. Minneapolis: Sage Publications, 2003.

———. "Reasonable Utility Functions and Playing the Cooperative Way." *Critical Review of International Social and Political Philosophy*, vol. 11 (2008): 215-234.

———. "The Place of Religious Belief in Public Reason Liberalism." in *Multiculturalism and Moral Conflict*, edited by Maria Dimovia-Cookson and P. M. R. Stirk. Oxford: Routledge: 19-37.

———. *The Order of Public Reason*. New York: Cambridge University Press, 2011.

Gaus, Gerald and Kevin Vallier. "The Roles of Religious Conviction in a Publicly Justified Polity: The Implications of Convergence, Asymmetry and Political Institutions." *Philosophy and Social Criticism*, vol. 35 (2009): 51-76.

Gauthier, David. *Morals by Agreement*. New York: Oxford University Press, 1986.

Goldman, Alvin. "Reliabilism." *Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*, 2008. Retrieved October 13th, 2010. From <http://plato.stanford.edu/entries/reliabilism/>.

Greenawalt, Kenneth. *Religious Convictions and Political Choice*. Oxford: Oxford University Press, 1988.

———. "Grounds for Political Judgment: The Status of Personal Experience and the Autonomy and Generality of Principles of Restraint." *San Diego Law Review*, vol. 30 (1993): 647-675.

———. *Private Consciences and Public Reasons*. New York: Oxford University Press, 1995.

Greene, Abner. "Uncommon Ground: A Review of Political Liberalism by John Rawls and Life's Dominion by Ronald Dworkin." *George Washington Law Review*, vol. 62 (1994): 646-673.

Gutmann, Amy and Dennis Thompson. *Democracy and Disagreement*. Cambridge: Belknap Press of Harvard University Press, 1996.

Habermas, Jurgen. *Legitimation Crisis*. Boston: Beacon Press, 1975.

———. "Reconciliation through the Public Use of Reason: Remarks on John Rawls's Political Liberalism." *The Journal of Philosophy*, vol. 92 (1995): 109-131.

———. "Religion in the Public Square." *European Journal of Philosophy*, vol. 14 (1996): 1-25.

———. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. New York: MIT Press, 1998.

———. *Moral Consciousness and Communicative Action*. Cambridge: MIT Press, 1999.

Harman, Gilbert. *Change in View: Principles of Reasoning*. Cambridge: The MIT Press, 1986.

Hauerwas, Stanley. *After Christendom? How The Church Is To Behave If Freedom, Justice, and a Christian Nation Are Bad Ideas*. Nashville: Abingdon Press, 1999.

Hayek, F. A. "The Use of Knowledge in Society." *American Economic Review*, vol. 35 (1945): 519-530.

———. *Law Legislation and Liberty, Volume 1: Rules and Order*. Chicago: University of Chicago Press, 1973.

Hobbes, Thomas. *Leviathan*, edited by Edwin Curley. Indianapolis: Hackett, 1994.

Horton, John. "Reasonable Disagreement." in *Multiculturalism and Moral Conflict*, edited by Maria Dimovia-Cookson and P. M. R. Stirk. New York: Routledge: 58-74.

James, Aaron. "Constructing Justice for Existing Practice: Rawls and the Status Quo." *Philosophy and Public Affairs*, vol. 33 (2005): 281-316.

- Joyce, George Howard. *Principles of Natural Theology*. New York: Longmans, Green and Co., 1922.
- Korsgaard, Christine. *The Sources of Normativity*. New York: Cambridge University Press, 1996.
- Kretzmann. *The Metaphysics of Theism: Aquinas' Natural Theology in Summa Contra Gentiles I*. Oxford: Clarendon Press, 1997.
- Larmore, Charles. *Patterns of Moral Complexity*. New York, NY: Cambridge University Press, 1987.
- Locke, John. *Two Treatises of Government and A Letter Concerning Toleration*, edited by Ian Shapiro. New Haven: Yale University Press, 2003.
- Locke, John and Peter H. Nidditch. *An Essay Concerning Human Understanding*. New York: Oxford University Press, 1975.
- Locke, John and I. Ramsey. *The Reasonableness of Christianity with A Discourse of Miracles and part of A Third Letter Concerning Toleration*. Stanford: Stanford University Press, 1958.
- Lomasky, Loren. *Persons, Rights, and the Moral Community*. New York: Oxford University Press, 1987.
- Macedo, Stephen. "In Defense of Liberal Public Reason: Are Slavery and Abortion Hard Cases?" in *Natural Law and Public Reason*, edited by Robert P. George and Christopher Wolfe. Washington, D.C.: Georgetown University Press: 11-49.
- (2011). Why Public Reason? Common Knowledge and Democratic Justice: forthcoming.
- MacIntyre, Alasdair. *Three Rival Versions of Moral Inquiry: Encyclopaedia, Genealogy, and Tradition*. Notre Dame: University of Notre Dame Press, 1990.
- Martin, Rex. *A System of Rights*. New York: Oxford University Press, 1993.
- McCarthy, Thomas. "Kantian Constructivism and Reconstructivism: Rawls and Habermas in Dialogue." *Ethics*, vol. 105 (1994): 44-63.
- McMahon, Christopher. *Collective Rationality and Collective Reasoning*. Cambridge: Cambridge University Press, 2001.
- Milbank, John. *Theology and Social Theory: Beyond Secular Reason*. Malden: Blackwell, 2006.
- Mill, John Stuart. *On Liberty*, edited by Elizabeth Rapaport. Indianapolis: Hackett Publishing Company, 1978.
- Mises, Ludwig von. *The Theory of Money and Credit*. New Haven: Yale University Press, 1953.
- Mize, Sandra Y. "The Common Sense Argument for Papal Infallibility." *Theological Studies*, vol. 57 (1996): 252-263.

Moreland, J. P. *Consciousness and the Existence of God: A Theistic Argument*. London: Routledge Press, 2009.

Muldoon, Ryan. *Diversity and the Social Contract*: unpublished dissertation, 2009.

Nagel, Thomas. *The View From Nowhere*. New York: Oxford University Press, 1986.

———. "Moral Conflict and Political Legitimacy." *Philosophy and Public Affairs*, vol. 16 (1987): 215-240.

Neal, Patrick. "Religion within the Limits of Liberalism Alone?" *Journal of Church and State*, vol. 39 (1997): 697-722.

———. "Is Political Liberalism Hostile to Religion?" in *Reflections on Rawls: An Assessment of His Legacy*, edited by Shaun P. Young. Burlington: Ashgate: 153-176.

Noddings, Nel. *Caring: A Feminine Approach to Ethics and Moral Education*. Los Angeles: University of California Press, 2003.

Nozick, Robert. *Anarchy, State and Utopia*. New York: Basic Books, 1974.

Pappas, George. "Internalist vs. Externalist Conceptions of Epistemic Justification." *Stanford Encyclopedia of Philosophy*, 2005. Retrieved October 13th, 2010. From <http://plato.stanford.edu/entries/justep-intext/#3>.

Perry, Michael. *Morality, Politics and Law*. New York: Oxford University Press, 1988.

———. *Love and Power: The Role of Religion and Morality in American Politics*. New York: Oxford University Press, 1991.

———. "Religious Morality and Political Choice: Further Thoughts--And Second Thoughts--On Love and Power." *San Diego Law Review*, vol. 30 (1993): 703-727.

Pettit, Philip. *The Common Mind: An Essay on Psychology, Society and Politics*. New York: Oxford University Press, 1996.

Pincione, Guido and Fernando Tesón. *Rational Choice and Democratic Deliberation: A Theory of Discourse Failure*. New York: Cambridge University Press, 2006.

Plantinga, Alvin. *God and Other Minds: A Study of the Rational Justification of Belief in God*. Ithaca: Cornell University Press, 1990.

———. *Warrant and Proper Function*. New York: Oxford University Press, 1993.

———. *Warranted Christian Belief*. New York: Oxford University Press, 2000.

Pollock, John. *Thinking about Acting*. New York: Oxford University Press, 2006.

Pollock, John and Joseph Cruz. *Contemporary Theories of Knowledge*. Lanham: Rowman and Littlefield, 1986.

Quong, Jonathan. "The Scope of Public Reason." *Political Studies*, vol. 52 (2004): 233-250.

———. *Liberalism without Perfection*. New York: Oxford University Press, 2010.

Railton, Peter. *Facts, Values, and Norms: Essays Towards a Morality of Consequence*. Cambridge: Cambridge University Press, 2003.

Rawls, John. "Outline of a Decision Procedure for Ethics." *The Philosophical Review*, vol. 60 (1951): 177-197.

———. "Justice as Fairness." *The Journal of Philosophy*, vol. 54 (1957): 653-662.

———. "Justice as Fairness." *The Philosophical Review*, vol. 67 (1958): 164-194.

———. *A Theory of Justice*. New York: Oxford University Press, 1971.

———. "Political Liberalism: Reply to Habermas." *Journal of Philosophy*, vol. 92 (1995): 132-180.

———. "The Idea of Public Reason Revisited." *The University of Chicago Law Review*, vol. 64 (1997): 765-807.

———. *Justice as Fairness: A Restatement*, edited by Erin Kelly. New Delhi: Universal Law Publishing Co. Pvt. Ltd., 2001.

———. *Political Liberalism*. New York: Columbia University Press, 2005.

———. *Lectures on the History of Political Philosophy*, edited by Samuel Freeman. Cambridge: Belknap Press, 2007.

Raz, Joseph. *The Morality of Freedom*. Oxford: Oxford University Press, 1986.

Reid, Thomas. "Essay on the Intellectual Powers." in *Thomas Reid's Inquiry and Essays*, edited by Ronald E. Beanblossom and Keith Lehrer. Boston: Hackett Publishing Company: 127-296.

Reidy, David. "Rawls's Wide View of Public Reason: Not Wide Enough." *Res Publica*, vol. 6 (2000): 49-72.

Ridge, Michael. "Contractualism and the New and Improved Redundancy Objection." *Analysis*, vol. 63 (2003): 337-342.

Rorty, Richard. "Religion as a Conversation-Stopper." in *The Ethics of Citizenship: Liberal Democracy and Religious Convictions*, edited by J. Caleb Clanton. Waco: Baylor University Press.

Rousseau, Jean-Jacques. *Rousseau: 'The Social Contract' and Other Later Political Writings*, edited by Victor Gourevitch. New York: Cambridge University Press, 1997.

- Sandel, Michael. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press, 1982.
- Scanlon, Thomas. *What We Owe To Each Other*. Cambridge: Harvard University Press, 1998.
- . *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge: Harvard University Press, 2008.
- Schwartzman, Micah. "The Completeness of Public Reason." *Politics, Philosophy and Economics*, vol. 3 (2004): 191-220.
- . "The Sincerity of Public Reason." *Journal of Political Philosophy*, vol. (2011): forthcoming.
- Searle, John. *Intentionality: An Essay in the Philosophy of Mind*. New York: Cambridge University Press, 1983.
- . *The Rediscovery of the Mind*. Cambridge: MIT Press, 1992.
- Sidgwick, Henry. *The Methods of Ethics*. London: Macmillan, 1907.
- Stich, Stephen. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge: MIT Press, 1990.
- Strawson, P. F. *Freedom and Resentment and Other Essays*. New York: Routledge, 1974.
- Suddoth, Michael. "Defeaters in Epistemology." *Internet Encyclopedia of Philosophy*, 2008. Retrieved November 8th, 2010. From <http://www.iep.utm.edu/ep-defea/>.
- Swinburne, Richard. *The Evolution of the Soul*. New York: Oxford University Press, 1997.
- . *The Existence of God*. Oxford: Oxford University Press, 2004.
- . *Revelation: From Metaphor to Analogy*. New York: Oxford University Press, 2007.
- Taylor, Charles. *Sources of the Self: The Making of the Modern Identity*. Cambridge: Cambridge University Press, 1989.
- Thomson, Judith Jarvis. "A Defense of Abortion." *Philosophy and Public Affairs*, vol. 1 (1971): 47-66.
- Tomasi, John. *Liberalism beyond Justice: Citizens, Society, and the Boundaries of Liberal Political Theory*. Princeton: Princeton University Press, 2001.
- Tutu, Desmond. *No Future Without Forgiveness*. New York: Doubleday, 1999.
- Wall, Steven. *Liberalism, Perfectionism and Restraint*. New York: Cambridge University Press, 1998.
- . "On Justificatory Liberalism." *Politics, Philosophy and Economics*, vol. 9 (2010): 123-149.

Weithman, Paul. *Religion and the Obligations of Citizenship*. New York: Cambridge University Press, 2002.

Wenar, Leif. October 12th. "Rights." *Stanford Encyclopedia of Philosophy*, 2010. Retrieved October 13th, 2010. From <http://plato.stanford.edu/entries/rights/>.

Wilde, Oscar. *The Picture of Dorian Gray*. Hertfordshire: Wordsworth Editions Limited, 1992.

Williams, Bernard. *Moral Luck*. Cambridge: Cambridge University Press, 1981.

Williams, Roger and James Calvin Davis. *On Religious Liberty: Selections from the Works of Roger Williams*. Cambridge: The Belknap Press of Harvard University Press, 2008.

Wittgenstein, Ludwig. *On Certainty*. Malden: Blackwell, 2003.

Wolf, Susan. *Freedom within Reason*. Oxford: Oxford University Press, 1990.

Wolterstorff, Nicholas. "The Role of Religion in Decision and Discussion of Political Issues." in *Religion in the Public Square: The Place of Religious Convictions in Political Debate*. Lanham: Rowman and Littlefield: 67-120.

———. "Why We Should Reject What Liberalism Tells Us about Speaking and Acting in Public for Religious Reasons." in *Religion and Contemporary Liberalism*, edited by Paul Weithman. Notre Dame, IN: University of Notre Dame Press: 162-181.

———. "The Paradoxical Role of Coercion in the Theory of Political Liberalism." *Journal of Law, Philosophy and Culture*, vol. 1 (2007): 101-125.

———. *Practices of Belief: Volume 2, Selected Essays*, edited by Terence Cuneo. New York: Cambridge University Press, 2009.

Yates, Melissa. "Rawls and Habermas on Religion in the Public Sphere." *Philosophy and Social Criticism*, vol. 33 (2007): 880-891.

Zagorin, Perez. *How the Idea of Religious Toleration Came to the West*. Princeton: Princeton University Press, 2003.