PLURALISM AND STABILITY:
A NEW APPROACH TO RELIGIOUS ACCOMMODATION

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Government

By

David Golemboski, M.T.S.

Washington, D.C.
June 22, 2016
PLURALISM & STABILITY:
A NEW APPROACH TO RELIGIOUS ACCOMMODATION

David Golemboski, M.T.S.

Thesis Advisor: Richard A. Boyd, Ph.D.

ABSTRACT

Freedom to hold and act on one’s deepest commitments is a basic liberal commitment, but religious beliefs routinely conflict with the requirements of the law. The question then arises: when (if ever) do sincerely-held religious beliefs warrant a dispensation from legal obligations? Legal and political philosophers alike have struggled to formulate a coherent and satisfactory resolution to this challenge presented by pluralism, generally theorizing exemptions as a matter of equality. I explore an alternative approach that has not been extensively theorized, despite the centrality of its basic concern: stability. I argue that the political goal of stability can shed light on both the challenges and opportunities presented by pluralism, and can provide normative guidance for addressing conflicts in practice. Moreover, the stability-based framework I defend is largely consistent with core liberal commitments. I argue that stability supports some exemptions to relieve citizens of religious burdens and to support the flourishing of diverse moral communities, which, I argue, are indirectly supportive of stability. I engage with legal scholarship and American jurisprudence on religious accommodation to provide novel takes on past and present cases.
ACKNOWLEDGEMENTS

The first people who deserve thanks upon the completion of this project are the four members of my dissertation committee. Richard Boyd has supplied precisely the right mix of encouragement, advice, and latitude as I’ve written. I have learned from Richard a great deal about what it means to be a political theorist, and I will do well to imitate his example as a scholar, teacher, and mentor. Bruce Douglass has been helpful and encouraging since the moment I arrived at Georgetown. He has been a perceptive reader and a lively conversation partner. Henry Richardson has been more than generous in joining this committee and in applying his keen critical eye to successive drafts of each chapter. He has helped me improve crucial components of the argument, and his imprint on the text as a whole is deep. Finally, Michael Kessler has as much to do with the genesis of this project as anyone, for it was in a seminar of his that I first wrestled seriously with the problem of religious accommodation. He has been gracious in supporting this effort, and I take inspiration from his example of multidisciplinary training and scholarship.

The Government Department at Georgetown has been an eminently satisfying setting in which to carry out this work. I am grateful for a generous department fellowship, as well as other funding, including a summer research grant to support this dissertation in 2015. Faculty apart from my committee have been encouraging and supportive, and a few graduate student colleagues deserve mention as well: Sean Coleman was a much-appreciated companion as we stumbled into the world of political science five years ago – and he has been much missed since he left for other things; Gianna Englert has been a model colleague and a friendly, refreshing presence in the department.
This dissertation is unquestionably better for my friendships with Nick McDaniel and Dan Horan, even though I don’t think I have discussed its contents at length with either of them. They have each contributed immensely to the sharpening of my mind through countless arguments – always well-populated with laughter. With apologies for lumping them together, I am deeply grateful for each of them.

My final and most exuberant thanks go to my wife, Brianna Copley. For supporting me – materially and otherwise – while I’ve indulged myself in this degree, she has earned my profound gratitude. For doing so while engaged daily in the genuinely heroic work of educating young people who are among the most disadvantaged in the country, she has earned my sincere admiration. Her enthusiasm and encouragement on this dissertation, while invaluable, account for but a drop of water in the comprehensive ocean of gladness that she has brought to my life.
# Table of Contents

Introduction: Liberalism, Pluralism, and the Search for Stability ................................................ 1

0.1 Pluralism: Peril or Promise for Liberalism ........................................................................... 1

0.2 The Problem of Religious Exemptions .............................................................................. 7

0.3 Motivating Ends: The Search for Stability ....................................................................... 11

0.4 Liberal Impartiality: Between Pluralism and Principles ............................................... 15

0.5 Methodological Notes ....................................................................................................... 18

0.6 Plan of the Dissertation ..................................................................................................... 25

0.7 What is Impartiality? ........................................................................................................... 29

Chapter One: The Egalitarian Basis of Impartiality ................................................................. 44

1.1 Moral Sentiments: Smith’s Impartial Spectator ............................................................... 46

1.2 Utilitarianism: Equal Regard for Interest ......................................................................... 51

1.3 Liberal Rationalism: Rawls’s Original Position and the Idea of Public Reason ........... 55

1.4 A Generic Argument from Equality to Impartiality ....................................................... 62

1.5 The Indeterminacy of Equality: Religious Exemptions ............................................... 74

1.6 Conclusion .......................................................................................................................... 91

Chapter Two: Stability for the Right Reasons ........................................................................ 94

2.1 The Case for a Modus Vivendi ............................................................................................ 97

2.2 Two Problems: Free-Riding and Mutual Assurance ....................................................... 100

2.3 Stability for the Right Reasons ........................................................................................ 104

2.4 Pluralism and the Reasonableness of SFRR ................................................................. 112

2.5 Stability and Coercion ....................................................................................................... 121
2.6 The Politics of Stability ................................................................. 125
2.7 Why Stability in the First Place? .................................................. 128
2.8 Situating Stability in Ideal and Non-Ideal Theory ................................ 134
2.9 Approximating Justice .................................................................. 138

Chapter Three: From Stability to Impartiality: A Hobbesian Account .............. 143
3.1 Fear and Hope .............................................................................. 145
3.2 Hobbesian Stability ...................................................................... 150
3.3 Stability and the “Perils of Pluralism” ............................................. 156
3.4 Impartiality in Hobbes ................................................................. 169
3.5 Stability and Distributive Impartiality ............................................ 181
3.6 Stability and Justificatory Impartiality ............................................ 194

Chapter Four: Overcoming Impartiality’s Motivational Deficit: Some Approaches .......... 202
4.1 Impartiality’s Motivational Deficit: The Problem of Ineffective Endorsement .. 203
4.2 Civic Identity .................................................................................. 211
4.3 Political Emotions ......................................................................... 229
4.4 Harmonization of Loyalties ............................................................ 240
4.5 The Promise of Pluralism ............................................................... 249
4.6 Moral Communities as a Source of Stability .................................... 258

Chapter Five: Stability and the Accommodation of Religious Belief ......................... 264
5.1 The Ambivalence of Stability ......................................................... 266
5.2 Some Guiding Principles .............................................................. 279
5.3 U.S. History and the RFRA Balancing Test ..................................... 307
5.4 The Case of Hobby Lobby ............................................................... 312
5.5 A New Test: Comparing the Stability and RFRA Frameworks ........................................... 326

5.6 Some Challenges ........................................................................................................... 330

Conclusion: Assessing the Stability Approach .................................................................... 344

6.1 Recapitulation ................................................................................................................ 344

6.2 Some Elements of Significance ..................................................................................... 348

6.3 Caveats and Limitations ............................................................................................... 352

6.4 Future Trajectories ....................................................................................................... 356

6.5 Religion and Liberal Politics ......................................................................................... 363

Bibliography ......................................................................................................................... 369
INTRODUCTION:

LIBERALISM, PLURALISM, AND THE SEARCH FOR STABILITY

0.1 Pluralism: Peril or Promise for Liberalism

The principal commitment of liberal politics is the guarantee of wide latitude for individuals to live as they wish. From Mill’s harm principle to Rawls’s lexical prioritizing of maximum individual liberty, the hallmark of liberalism has been the freedom of citizens to live as they wish. Though, this guarantee also entails reciprocal limitations: an individual’s liberty extends only so far as is consistent with the interests of others (Mill) or the enjoyment of a similar scheme of liberties by other (Rawls). The exercise of liberty is always bounded by the legitimate claims of one’s fellows, and the project of liberal political philosophy has been to identify these bounds conceptually.

This project is complicated by the fact of pluralism, by which I refer to the proliferation of different conceptions of the good within a single political society. It is worth taking a moment to distinguish this usage from some others. Consider three different ways in which the term ‘pluralism’ may be invoked.

The first describes a normative political view that calls for the recognition and maintenance of diverse social entities (of which the state is but one), each possessing their own distinctive dignity and authority, within a single political society. Proponents of this kind of normative pluralism include British thinkers such as John Neville Figgis, G.D.H. Cole, and Harold Laski.¹ They take the “foundational autonomy”² of diverse associations as a starting

point, and call for political institutions that defer appropriately to these associations under a common legal order. A second variety of pluralism is the value pluralism associated with authors such as Isaiah Berlin, Joseph Raz, and William Galston. This view takes off from the premise that there are multiple irreducible goods which cannot be definitively ranked against one another. Value pluralism is opposed to value monism; these are meta-ethical positions, not strictly political ones. Proponents of value pluralism do, however, tend to favor political arrangements that grant significant latitude for individuals to affirm and live in accordance with diverse value systems. A third usage of ‘pluralism’ is strictly descriptive: it signifies the fact that persons affirm many diverse beliefs, maintain many diverse commitments, and pursue many diverse ways of life. We might call this sociological pluralism, in that it merely denotes the variety present in some given set of people. It is in this sense that a great many authors talk of pluralism. Rawls is one example, though he further specifies that liberal societies give rise not only to pluralism, but to reasonable pluralism, by which he refers to pluralism arising from the free exercise of human reason, as opposed to class interests or individuals’ limited political imaginations. This distinction noted, it remains that what Rawls means by pluralism is the simple descriptive fact of disagreement.

My project here has something to do with each of these usages. Most basically, when I say that the fact of pluralism complicates the determination of liberty’s proper bounds, I am referring to pluralism in the sociological sense. Specifically, I am focused on the fact that individuals affirm and pursue diverse conceptions of the good. Unless I indicate otherwise, the

---

descriptive fact is what I refer by ‘pluralism.’ In chapter 3, I take onboard Rawls’s (and Hobbes’s) conviction that this diversity is not an accident of circumstance or faulty reason, but rather an inevitable consequence of human experience. I do not take a position on whether the sociological fact of pluralism derives from objective value pluralism. Disagreement is a clear and present political reality regardless of whether it stems from diverse interpretations of a single *summum bonum* or from diverse ways of reconciling multiple incommensurable goods. That said, it is the case that many people experience morality in terms of value pluralism, as a matter of balancing competing goods, desires, or obligations. From a phenomenological vantage, value pluralism may as well be true. That said, nothing in what follows depends on the objective truth of the metaethical claims of value pluralism. Finally, while I do not begin my argument with any independent commitment to a pluralism of associations, a la the British pluralists, I do end up affirming that the proliferation of social entities is a good thing, in some important respects.

The fact of pluralism complicates the project of liberal politics. One reason for this is that citizens will desire to utilize their liberties for diverse purposes. Much of the time, this is politically insignificant: one person wants to buy coffee, another tea; one person wants to worship at an Episcopal church, another at an orthodox synagogue. But differences in aims, projects, and preferences can sometimes lead to asymmetries of liberty. Some people will be allowed to pursue the ends they wish while others will not. Persons who prefer steak and chicken, respectively, are equally free to consume their preferred meat, but persons who prefer wine and heroin, respectively, are not equally free to consume their preferred mind-altering substance. This is a mundane example, but it points up the fact that not all choices are equally permissible. Pluralism means that the matrix of goods that a public will pursue are not uniform,
and that extent of each person’s liberty to pursue the life she chooses will vary in relation to each’s profile of desires.

The other way in which pluralism complicates liberal politics is better characterized in terms of obligation than in terms of choice or preference. Pluralism means that people will understand themselves to bear different sets of moral obligations. Because individuals do not share a conception of the good, the duties that they believe apply to themselves will not be perfectly overlapping. Some people, for instance, will feel obligated to serve their nation in military service, while others will feel obligated to refuse to commit acts of violence. Some persons will feel duty-bound to observe certain norms of dress, while others will regard their clothing choices as morally indifferent. Some people will feel obligated to refrain from certain activities – using electricity, fighting in wars, receiving medical interventions, working on the Sabbath, to name a few – while others will regard these as morally permissible. What this means is that there will inevitably be some laws that conflict with some citizens’ moral obligations but not with others’. This is a challenge for liberal politics in that political obligations may place some people, but not others, in a conflict between competing obligations.

These challenges are real, but just how significant they are is less easy to determine. After all, it is obvious that individual liberties must be limited within political society: the priority of airline safety means that one cannot carry a knife onboard. Similarly, moral obligations will sometimes conflict with requirements of the law: witnesses in criminal trials are regularly called on to testify against persons toward whom they feel obligations of loyalty. As Rawls writes (channeling Berlin), “there is no social world without loss; that is, no social world that does not exclude some ways of life that realize in special ways certain fundamental values.” This would be the case even in a situation of complete social homogeneity – i.e. the absence of
pluralism. But the dynamics are complicated further by pluralism. While everyone in political society will have some of their desires limited, pluralism means that some people may have more of their desires frustrated than others. While everyone will have to navigate conflicts between their conception of morality and the requirements of the law, pluralism means that some people may have to face more of these conflicts than others.

Pluralism is not only a problem, however. Nor is it some accidental feature of liberal politics. It is, rather, embedded in the very idea of a free society. J.S. Mill recognized that one consequence – indeed, the very reason for which liberty is to be valued – is that free citizens will choose to live in diverse ways. More recently, John Rawls recognized that the well-ordered society governed by liberal norms of justice as fairness, as he described in A Theory of Justice, would encourage citizens to embrace a pluralism of comprehensive doctrines. It was this recognition that prompted him to revise parts of his theory in the papers that became Political Liberalism. Diversity of beliefs and ways of life is both the keystone virtue and the persistent albatross of liberalism. Some authors note liberalism’s unique ability to “show us how to live together in societies that harbor many ways of life,” while others point to the “perils” of pluralism, describing it as something that must be “survived.”

This ambivalence toward pluralism has marked liberal discourse throughout its history and continuing to the present day. The diverse approaches that liberal thinkers have taken to pluralism might be categorized in any number of ways, but one way to characterize the

---

intellectual landscape is in terms of two distinct strains of thought. Republican liberalism, exemplified by such figures as Tocqueville and Rousseau, has advocated civic solidarity through shared identity that, while not entirely suppressive of pluralism, binds citizens together through a core of shared identity and principled commitment. Another approach, what might be called individual rights liberalism, seen in figures such as Locke and Mill, has more emphatically endorsed pluralism as a matter of personal choice, and has tried to draw limits on the basis of reciprocal negative right claims. In contemporary political philosophy, proponents of political liberalism clash with liberal perfectionists, the former affirming neutrality among conceptions of the good, and the latter arguing that political orders can and should promote certain liberal ideals. The ambivalence of liberals extends to treatment of not only diverse citizens, but the diverse groupings into which free citizens sort themselves, with liberals being historically divided over how to regard associations, and how much latitude to afford them. Liberals, in agreement on the basic priority of choice and individual freedom, have always disagreed on location of liberty’s bounds, and the principles or values that ought to ground the determination of those bounds. At a very basic level, the following question has dangled over liberalism in each of its historical and conceptual formulations: is pluralism a welcome sign of liberalism’s virtues, or a threat to the coherence and integrity of liberal order? Is difference to be celebrated and affirmed, or merely tolerated, and always with a suspicious eye?

0.2 The Problem of Religious Exemptions

One front in the conflict of liberties arising from pluralism concerns the right of religious exercise: the freedom to hold and act on a set of religious beliefs, including the moral commitments those beliefs include or entail. Religious liberty has long occupied a central, if not an outright privileged, position in liberal thought. Indeed, it is often referred to as the “first” freedom, and not only because it is the first right enumerated in the Bill of Rights. Liberal political orders tend to codify protection for religious exercise in their basic legal frameworks, as in the protection afforded by the First Amendment to the U.S. Constitution. But a legal grant of religious liberty inevitably runs into the following problem: religious beliefs routinely conflict with the obligations imposed by the law. That is, a legal grant of religious liberty may create conflicts for the law itself, as citizens will have at least a minimally legitimate claim to carry out their religious practices even when they violate other legal obligations.

Note two things about the claim I have just made. First, I asserted that religious beliefs can conflict with legal obligations. This may seem like a category mistake, as beliefs are matters of cognition while obligations are matters of action – they operate in different domains, so to speak of ‘conflict’ may be entirely mistaken. What I mean, of course, is that religious beliefs often entail practical norms that amount, in some cases, to obligations – and that these

---


14 A brief note to specify what I mean by ‘exemption.’ I mean to describe those dispensations from standing legal obligations that do not imply an alteration of the content or general applicability of those obligations. This is to distinguish phenomena like the following two: 1) a legislature creates an exemption from employment law for those who refuse to work on a given day because they are celebrating their Sabbath; and 2) a legislature crafts the employment law to hold that persons may never be fired for refusing to work on a given day of the week. In the former case, the requirement to work on any day of the week is lifted only in virtue of a person’s religious observance of a Sabbath; in the latter, the requirement to work on any day of the week is eliminated entirely, for everyone. Thus, the former qualifies as an exemption, while the latter is more properly described as a change in the baseline law. An exemption is always a dispensation from an obligation in virtue of some attribute/commitment/etc. This will become clearer following the discussion of impartiality below.
obligations can conflict with legal obligations. For example, the belief that killing is always and everywhere wrong entails a practical obligation to abstain from killing.\textsuperscript{15} This might conflict with a citizen’s legal requirement to fight in a war. The same dynamic obtains whether religious and legal obligations are negative or positive, forbidding action or requiring it. Saying that religious “beliefs” may conflict with legal obligation is shorthand for the ways in which beliefs can motivate religious obligations. These obligations are not optional components of a religious existence; as Rob Vischer notes: “a fully integrated life is not possible unless the dictates of conscience are reflected in action.”\textsuperscript{16} I will also, in what follows, speak at times of religious “commitments,” which further expands the set of actions that have their basis in religion, though they may not take the form of obligations. For instance, that clergy must wear certain garments during ritual ceremonies may not be, strictly speaking, an obligation, but it is nevertheless an important part of religious exercise.

The second thing to note is my claim that conflict between religious belief and legal obligation is routine. It would be impossible to substantiate this claim in a definitive empirical way, for such conflicts are ordinarily reserved at the level of the individual believer, and leave no formal record. But the claim should not come as a surprise: insofar as religious beliefs bear on actions, those actions are susceptible to regulation by laws. Moreover, insofar as laws affirm or reflect values,\textsuperscript{17} these values may conflict with some citizens’ religious beliefs. I won’t belabor the point here, but I trust that the potential for religious-legal conflict is obvious. The “two

\textsuperscript{15} I am assuming that it is incoherent to affirm both the proposition “X is immoral” and “I have no obligation to abstain from X.” This may be questionable, but in ordinary cases it accurately describes the way moral beliefs inform practical obligations.


“kingdoms” subject to religious and political authority are, if not one and the same, at the very least substantially overlapping.

To the extent that these conflicts appear to arise only infrequently, I would suggest that this is not because conflict between religious and legal obligation actually arises infrequently, but rather because such conflict is so eminently ordinary that we lose sight of the legal mechanisms we have developed for handling it smoothly and noncontroversially. Take, for instance, the fact that the Roman Catholic Church does not, as a matter of doctrine and policy, ordain women as priests; consequently, employment opportunities that must be filled by priests are not open to women.\footnote{A small number of Catholic parishes have hired women to serve as lay pastoral administrators, performing many functions that priests would ordinarily cover. Core sacramental functions, however, remain the sole province of ordained men. Ruth A. Wallace, \textit{They Call Her Pastor: A New Role for Catholic Women} (Albany: State University of New York Press, 1992); James Coriden, “The Changing Structures of Parish Ministry,” \textit{New Theology Review} 22 (2009): 77–79.} In general, of course, employers in the United States are forbidden by Title VII of the Civil Rights Act from discriminating on the basis of sex. American courts, however, have since the passage of the Act affirmed a “ministerial exception,” grounded in the First Amendment, which precludes the application of antidiscrimination legislation to religious institutions’ selection of their ministers. The bounds of this exception have at times been controversial,\footnote{Hosanna-Tabor Evangelical v. Equal Employment Opportunity Commission, 565 U.S. \underline{___} (2011).} but for the most part its implications are accepted by Americans as a matter of course. This is an example of the degree to which accommodation of religion when it conflicts with legal obligation is embedded in liberal political culture. But this is not evidence that the conflict between religion is a marginal one; to the contrary, it is evidence of its pervasiveness.

The rule-and-exemption framework is very often successful in mitigating conflicts arising from religious exercise, but it is no surprise that it also raises frequent controversy. After all, religious exemptions do not resolve the conflict between religious belief and legal obligation, but
rather dispatch with it by subordinating the latter to the former. Insofar as laws are designed to serve worthy ends approved by some legitimate democratic process, exempting some persons from those laws might reasonably raise concern. After all, compliance with the law is not generally considered optional, and democratic sensibilities chafe at the idea of some individuals receiving exceptional treatment. Nevertheless, a liberal regime that cherishes religious freedom ought also to hesitate before simply assuming the dominance of legal obligation when it comes into conflict with religious obligation. In recent years, some elements in American religious communities have waged a fevered push in support of religious liberty, much of which has focused on securing religious exemptions from laws that offend their religious beliefs. The most prominent of these was the case of *Burwell v. Hobby Lobby* (2014), in which the owners of a chain of craft stores requested an exemption from a requirement to provide their employees with health insurance that covered contraceptive drugs the use of which, they claimed, was prohibited by their religious beliefs.\(^{20}\)

Like all legal cases, *Hobby Lobby* turned on the particularities of the religious claim and the statutory specifics governing exemptions in the U.S. in 2014. But in its basic logic, *Hobby Lobby* dealt with the fundamental question raised by the idea of religious exemptions: When, if ever, does an individual or group’s religious belief or identity warrant a dispensation from legal obligations? Two categorical answers are possible, though undesirable. *Never* permitting exemptions would mean forcing the Catholic church to accept women as priests, insisting that religious pacifists fight in wars, and requiring clergy to give testimony in court detailing information they learned during sacramental confession. On the other hand, *always* granting religious exemption requests would mean that there would be no bounds to what a religious person could do in the name of her faith; human sacrifice is an oft-cited example of conduct that

---

simply cannot be permitted even if commanded by religion. The answer, then, must lie
somewhere between these two poles. Determining the degree of permissiveness or
restrictiveness that should be afforded religious exercise by means of exemptions amounts,
therefore, to wrestling with the basic challenge of pluralism, crystallized in the specific problem
of religious exemptions.

0.3 Motivating Ends: The Search for Stability

A principled answer to the question of religious exemptions must proceed on the basis of
some end or ends. Some value, good, or goal must supply a ground for norms which we can use
to formulate parameters, evaluate specific cases, and assess the merits of outcomes. (Even a
strictly “practical” approach relies implicitly on a guiding end, such as feasibility, conflict-
mitigation, or the like.) In this dissertation, I develop an approach based on the end of stability. I
take political stability to be a political good, and then ask what kinds of exemptions are
conducive to that end. Answering the question will divide exemptions into three categories: those
that are necessary for stability, those that are incompatible with stability, and those that are
permissible in a stable regime. My investigation (especially the discussion in Chapter 5) deals
principally with the first two categories, looking for the necessary conditions for stability.

In order to proceed, of course, we must have some conception of what we mean by
‘stability.’ Does stability mean the achievement of basic order, where people are, if nothing else,
prevented from killing one another? Or does stability mean the endurance of a set of laws and
institutions over time? Is a political order stable if some, but not all, of its laws endure over time?

Following Brettschneider, we might refer to these as dystopian possibilities. Though the proper balance is yet
unknown, it is clear enough that neither of these poles represents an acceptable approach. Corey Brettschneider,
When the State Speaks, What Should It Say? How Democracies Can Protect Expression and Promote Equality
Or if some, but not all, of its citizens refrain from violence toward one another? Moreover, we must ask whether any and all means to achieving stability are acceptable. A passive populace terrified into submission to some tyrant can be “stable,” as can a decadent and lazy people unwilling to stir themselves to disrupt the prevailing order. Do these scenarios qualify as achievements of stability? For the argument of this dissertation to proceed, we must establish a proper conception of stability that can serve as a normative ideal, and from which we can derive the principles that constitute necessary conditions for its achievement. Specifying this conception is the task of Chapter 2, where I argue that the conception of stability worthy of treatment as a normative ideal is one that grounds stability in citizens’ endorsement of their regime, and derives popular support on the basis of that endorsement, such that the basic structure of the order will endure over time. This rules out stability-by-tyranny, and suggests that stability requires the endurance of central elements, but not all details, of a political order. It is, I argue, appropriating Rawls’s notion, stability for the right reasons.22

To ask what kinds of political principles are required in order to achieve stability is, in some important ways, a retrieval of a basic liberal concern. Historically, liberal thought originated in contexts of conflict and instability, and sought political frameworks to mitigate these problems. Theories of religious toleration are the clearest example of this effort, as religious conflict was on the minds of many progenitors of the liberal tradition.23 A concern for stability is inherent in the social contract tradition, insofar as one goal of contractarianism is to address the coordination problems arising from social cooperation.24 Both Hobbes and Locke, at

22 Rawls, Political Liberalism, xl.
24 I do not mean to imply the strict identity of liberalism and contractarianism. Contract views can certainly be illiberal (see Hobbes), and liberal views need not rest on a supposed contract (see Mill). Yet, there is something central to the social contract idea that is also deeply central to the liberal tradition: the conception of citizens as free and equal. For this reason, I speak of the two as, if not exactly interchangeable, rather closely related.
least, understood this to be one of the functions of a social contract, and it can be seen as well in Rousseau and Rawls. Contract solutions provide an equilibrium from which citizens’ balance of reasons tilt in favor of maintaining their compliance with the political order. Contractarianism is not exclusively about coordination, however, as there is also a moral dimension: contract solutions reflect principles that are compatible with citizens’ freedom, in that citizens can (in some sense) agree to them or endorse them, and with equality, in that each citizen is regarded as exercising veto power over any potential agreement. Essentially, the contract tradition reflects a principle of respect for persons.\textsuperscript{25} Over the history of liberal thought, it is the moral dimension of contractarianism – and to liberal politics more generally – that has assumed the place of greatest prominence. One thing this project aims to do is reassert the centrality of stability as a (if not the) motivating concern for liberalism.

One way to cast this aim is in contrast to Rawls’s project throughout his work. The first two parts of \textit{A Theory of Justice} explicate and justify the conception of justice he called ‘justice as fairness,’ consisting of his famous two principles of justice. The third part of the book provides an argument for the stability of justice as fairness, attempting to show that citizens living in a well-ordered society governed by the two principles would come to reliably uphold those principles. Over time, Rawls came to doubt the stability argument in \textit{Theory}, and eventually revised that argument in \textit{Political Liberalism}.\textsuperscript{26} In both the early and later formulations, the stability argument served to establish the possibility of justice as fairness, to show that it did not describe an impossibly utopian ideal. It was only because of the moral desirability of the two principles that an investigation of their potential stability was warranted. Rawls’s goal was to specify the conditions for a society to be, as Joshua Cohen writes, “stably

\textsuperscript{26} Weithman, \textit{Why Political Liberalism}?
just,” with the adjective standing prior to the adverb. My aim, in contrast, is to reverse this order of priority: my primary end is mere stability, and once having established what stability requires in general, I ask what kinds of political principles stability implies. I ultimately show that a version of political liberalism (or at least its outlines) is a necessary condition for stability (of the proper kind). I therefore argue not only that stability deserves to be situated centrally among liberal concerns, but also that taking stability as a normative end can ground a good deal of the core commitments of liberalism.

I assume that the value of stability is largely self-evident. Political stability is objectively (though not necessarily unqualifiedly) a good thing. In a couple places throughout the dissertation, I make some brief comments in support of this claim, but these arguments are not extensive. The reason for this is that my objective in this project is not to establish the value of stability or its priority over against other political ends (such as justice or honor, for instance). Rather, I am interested to identify some of the necessary conditions – in the form of political principles – for stability. I want to know what kinds of political arrangements we ought to have in order for our political order to be stable. This leaves the argument open to an easy objection: that stability is not a sufficiently worthy end from which to derive political principles. (Because it is not itself very valuable or because there is some other, more valuable end that ought to order political arrangements.) In order for my investigation to get off the ground, however, we must assume that stability is an end from which it is appropriate to derive principles. I hope that the merits of this approach will be made known, in a sense, by its fruits: the coherence and plausibility of the principles it produces. Determining how much weight ought to be given to these conclusions and how they ought to be ranked against those derived from other ends are projects for another time.

0.4 Liberal Impartiality: Between Pluralism and Principles

Though it has not yet been mentioned, the concept of impartiality plays an important role in this project. Impartiality, I argue, is a core value in liberal political thought, and indeed a certain variety of liberalism can be captured by two principles of impartiality. Thomas Nagel has written that “ethics always has to deal with the conflict between the personal standpoint of the individual and some requirement of impartiality,” and, moreover, that “the clash between impartiality and the viewpoint of the individual is compounded when we move from personal ethics to political theory.”28 This is particularly the case in liberal theory, from early doctrines of toleration and the rule of law to more recent accounts of political liberalism and liberal neutrality. On the other hand, impartiality has been a prime target of feminist and communitarian objections to contemporary liberalism.29 That politics represents a key site of the encounter between the individual and the collective entails that impartiality and its attendant conflicts are, as Nagel suggests, ineliminable features of political discourse and thought.

This is not, of course, to say that there exists a consensus on the proper understanding and consequence of impartiality. Rather, the idea of impartiality is implicated – and disputed – in a number of key practical and theoretical contemporary political controversies. In the practical realm, questions of distribution often hinge on an understanding of impartiality and its requirements. Whether the allocation of university admission slots ought to be impartial with regard to race, for example.30 Even the bare meaning of the concept is far from obvious, and colloquial usages are not always precise.

My claim is that a basic version of political liberalism, as elaborated by such authors as Rawls, Larmore, and Nagel, can be captured by two principles of impartiality: distributive and justificatory. I will specify these in further detail below, but the first, distributive impartiality, suggests that only some characteristics of persons, and not others, ought to affect the political distribution of burdens and benefits among them. The specification of distributive impartiality determines whether considerations such as race, gender, family name, wealth, or religious identity should influence the way in which one is regarded by the political order. A norm of anti-racism is a clear instantiation of this principle: a person should be neither advantaged nor disadvantaged, vis-à-vis the political order, on the basis of their racial identity. Political liberalism, in general, presumes that personal characteristics are, in general, not relevant to one’s political standing, and should not determine one’s allotted benefits and burdens.31 It is a defeasible presumption, but establishes a burden of justification on any alteration of political burdens and benefits based on some specific characteristic shared by one or more persons. The second principle precludes political arrangements that depend on the truth of some particular comprehensive doctrine. This is the commitment underlying many doctrines of liberal “neutrality,” and marks the difference between political liberalism and comprehensive conceptions of liberalism.32 Together, the principles of distributive and justificatory impartiality stipulate that the political order ought to be “no respecter of persons,” and that its laws and institutions be justifiable apart from some particular religious system or conception of the human good.

31 This is the insight behind Rawls’s veil of ignorance, though he dismisses most personal characteristics as morally -- rather than politically -- arbitrary. John Rawls, A Theory of Justice, revised edition (Cambridge, MA: Harvard University Press, 1999), 63.
32 Larmore writes that “[T]he specific way in which liberalism distinguishes the public from the private is not the feature most characteristic of political liberalism. Instead, the distinctive liberal notion is that of the neutrality of the state.” Larmore, Patterns of Moral Complexity, 42.
The concept of impartiality plays an important role in this dissertation. It provides the normative connection between the end of stability and specific policy prescriptions. The normative conception of impartiality I will advance specifies which features of persons – and which differences among persons – merit recognition and potential accommodation by the political order. It specifies how diverse, competing conceptions of the good may (or may not) serve as justificatory support for political arrangements. Impartiality is removed by one level of abstraction from specific political principles, but it plays an important argument in determining what the content of those principles ought to be. In the case of religious exemptions, the basic question is whether religious belief is a personal characteristic that ought to influence the distribution of burdens and benefits (by altering some citizens’ legal obligations), or if citizens ought to be treated uniformly regardless of religion.

The prior question, though, is: why impartiality? Without identifying some basis for the principles of impartiality, we will not only lack justification for affirming them, but we will lack necessary resources to know how they ought to be applied. In particular, we will lack grounds on which to determine the specification of “politically relevant” differences among persons. We can only determine the scope of liberal impartiality by attending to its grounding. The predominant approach in contemporary liberal theory is to ground liberal impartiality in moral commitments, the most common being either equality or autonomy (or both). My project here is to provide an alternative ground for impartiality – namely, stability. I argue that the normative ideal of stability “for the right reasons” can ground both distributive and justificatory impartiality, and on the basis of that account we can derive principles for adjudicating religious

---

34 Larmore says that political liberalism must have a moral basis. “Political Liberalism,” Political Theory 18 (1990): 341.
exemption requests. My primary goal is not to reject egalitarian or libertarian accounts of impartiality, but rather to show the merits of an alternative, “stabilitarian” account, and then to apply that account to the problem of religious exemptions. One contribution of this argument is to show that the core commitments of political liberalism can be derived not from moral commitments, but rather from a non-moral normative ideal of stability.

0.5 Methodological Notes

A few notes on the methodology of this project are in order before delving into the substance of the argument.

Religion

This discussion began with some observations on the fact of pluralism in liberal political societies, but quickly narrowed in focus, concentrating on the challenges posed by religious pluralism. For almost all the remainder of this text, I focus specifically on religious diversity, rather than on pluralism more generally. This is a prudential move, made for the sake of simplicity. Focusing on a specific dimension of pluralism makes possible a deeper discussion than would otherwise be feasible in a reasonable amount of space. This should not be taken to imply that religion is somehow categorically distinct from other kinds of belief or commitment. Some authors do insist on a distinction of this sort; Brian Leiter, for instance, writes that religion is distinct from other kinds of belief by virtue of being “epistemically insulated” from evidentiary dispute.35 I do not assume any such distinction for the purposes of this argument. One reason for this is the basic difficulty of defining ‘religion’ in a manner that is both parsimonious

and capacious enough to contain all the various beliefs and practices commonly referred to as “religious.”  

The more important reason is that I am not convinced that, as far as stability is concerned, religion is categorically different from other kinds of commitments. Rather, religion crystallizes the features that matter from the vantage of stability. First, religious belief is a paradigm example of the non-shared commitments that set citizens apart from one another. There are many other such commitments, but religious ones offer a familiar and readily identifiable example. In a sense, religious pluralism captures the relevant issues arising from the general fact of pluralism, and offers a particularly ripe point of intervention for this project. Second, religious beliefs are among the most strongly-held and highly-valued such non-shared commitments. Persons are normally more strongly convicted about religion than about many other beliefs or commitments they hold, and they tend to value their religious commitments more highly than many others. 

What this means is that conflicts over religious belief have the potential to be, as is well known, passionate and volatile. Resolving religious conflict, or at least addressing it in some sensible way, is therefore an important priority for politics and political philosophy. This need not imply that religious belief is fundamentally distinct from other species of belief nor that other kinds of belief could not give rise to the same or analogous concerns vis-à-vis stability. Rather, the focus on religion is for the purpose of narrowing the scope of our inquiry and bringing some key political issues into focus.

---


37 Some citizens may, of course, be uncertain and/or lukewarm regarding their religious commitments, but these persons do not give rise to the relevant challenges to stability. For the sake of this project, it is strongly-held and highly-valued religion that is of interest.
In the final chapter, I will return to the question of religion’s distinctiveness. Having considered the issue of religious exemptions, we will turn to the question of whether the arguments supporting religious exemptions also support exemptions for analogous phenomena, such as commitments of “conscience,” or activities of non-religious associations. Until then, I bracket any discussion of the relationship of religious pluralism to pluralism more broadly.

Context

The fact of pluralism is, of course, not confined to any particular setting. As I suggest in a couple places, I am doubtful that it is practically possible for any political society to successfully avoid the fact of pluralism in an age of globalization, international commerce, and expanding communication media (if ever). Authoritarian regimes may, nevertheless, treat pluralism in the same way they deal with popular unrest or economic distress: as a problem to be eliminated. Liberal societies, on the other hand, are forced to reckon with pluralism, rather than simply overcome it. They expressly permit individuals to hold and pursue differing conceptions of the good, and it is under liberal institutions that pluralism is most likely to flourish. For the purposes of this project, then, I assume a liberal political context, and it is a liberal conception of impartiality that I wish to interrogate. That is not to say that I assume liberal political principles as part of my constructive argument. That would be to assume the very thing which I hope to demonstrate – namely, the compatibility of liberal principles with the end of stability. But I assume that the argument will be most applicable to liberal contexts, and most interesting to liberal people, insofar as these are the ones most likely to be alert to the nuances and challenges of pluralism. In a couple places, I note where the argument from stability might sanction political arrangements that are to some degree illiberal, but I don’t pursue these suggestions in depth, for
two reasons: First, it is my contention that stability generally requires liberal political principles, so degrees of illiberality will be only minor amendments to general liberal norms. Second, illiberal suppression of pluralism, whatever its merits vis-à-vis stability, is not so much an effort to deal with the challenges of pluralism as an effort to preempt those challenges. For our purposes, in contrast, we must confront pluralism head-on.

Among liberal polities, of course, experiences of and approaches to pluralism differ. The United States, France, and South Korea each face different dynamics of diversity, and each responds to pluralism in its own way. With this in mind, I must acknowledge that the argument here is oriented principally to the North American experience, in some important respects. First, the example cases that I discuss are drawn mostly from the U.S. context. A few come from Canada or Great Britain, but most are American cases. Second, any analysis of legal norms or rules attends almost exclusively to U.S. constitutional and statutory law and jurisprudence. This limitation is for the sake of simplicity, in that it bounds the set of materials to accommodate and situates the constructive argument in relation to a single, coherent legal framework. While I intend the constructive argument to be a non-contextual one (I do not specify ‘American stability’ or ‘pluralism in the United States’), both the specific conflicts it addresses and the legal framework it assumes are distinctively American. I must be open to the possibility that it is shaded in ways that derive from my own situated-ness. If the argument relies on unstated assumptions gleaned from the American context, this is a weakness of the argument, not one of its intended features.

Finally, I must note that the foregoing argument deals with pluralism, stability, and religious accommodation on only the level of domestic politics. This is where the problem of religious exemptions arises most immediately, though the broader issues are of course of global
relevance. To address these issues on the international level would be an entirely different project, so we will confine ourselves to the domestic sphere.

Aims and Sources

The primary orientation of this dissertation is constructive, and it is largely in the mode of analytic political philosophy in the Anglo-American tradition. By this I mean that my primary objective is to advance an original normative argument about pluralism, stability, and the accommodation of religious difference, and that the best way to do this is by close attention to the normative concepts involved and their logical relationships to one another. This constructive project is aided by a couple of subsidiary aims.

First, the project is informed by attentiveness to various approaches to religious accommodation across different historical periods. In particular, we will see how different approaches to religious exemptions over the course of U.S. history exemplify the uncertain status of religious accommodation in liberal politics. Second, and relatedly, the project involves interpretations of constitutional, statutory, and judicial approaches to religious accommodation. Third, the project has a significant exegetical component, insofar as the constructive argument is largely informed by important thinkers in the liberal tradition. Each of chapters 1-4 relies to a significant degree on critical readings of historical and/or contemporary work in liberal theory. Fourth, the project is informed by a sociological understanding of religious difference and the empirical consequences of accommodation. As I will suggest, a stability-based approach to religious exemptions must be attuned to the ways in which those exemptions are likely to affect individuals and alter their perception of how the political order benefits and burden them.

short, the dissertation involves a number of distinct modes of inquiry, and aims to shed light on the issue of religious accommodations from various angles. That said, each of these elements serves primarily to advance the constructive project, supplying conceptual resources and support for the normative argument I develop.

The constructive orientation of the project informs the way in which different kinds of sources are employed throughout. In addition to a good deal of contemporary philosophical literature, I do draw on texts from the history of political philosophy – especially Hobbes’s *Leviathan*, but also works by Locke and Rousseau. In approaching these texts, I proceed on the assumption that the purpose of reading texts in the history of political thought is the same as the purpose of reading contemporary writing in political philosophy: to glean conceptual resources that may be applied toward philosophical argument. I read these texts *instrumentally* – treating them not as ends in themselves, but repositories of ideas that may be of use in addressing the questions that I wish to address. This is not to suggest that attention to biographical and contextual facts about these texts is unnecessary; to the contrary, understanding the history of the texts may be important to appreciating them fully. But I have not attempted to survey comprehensively the scholarly literature pertaining to each of the historical figures I engage. I have rather drawn selectively on interpretive scholarship that illuminates salient features of these canonical works. After all, my aim is not to understand Hobbes (say) for the sake of understanding Hobbes; rather, it is to understand Hobbes for the sake of more insightfully engaging with the normative issues before us.

Jeremy Waldron exemplifies the approach to historical texts that I have adopted here. Waldron routinely engages historical authors with a deft combination of philosophical rigor and regard for the integrity of the text. His skill is prominently on display in his various treatments of
John Locke—on property, on toleration, and on equality. In each of these instances, Waldron engages in rigorous textual analysis along with serious consideration of contextual factors that contribute to understanding. But he also focuses the bulk of his energy on the content, structure, and merits of Locke’s various arguments. In the introduction to his analysis of Locke’s argument in *A Letter on Toleration*, Waldron writes: “To put it bluntly, I want to consider whether Locke’s case is worth anything as an argument which might dissuade someone here and now from actions of intolerance and persecution.” This is a method that appropriates the lessons of the historian and combines them with the ambitions of the philosopher. In *God, Locke, and Equality*, Waldron summarizes how these two perspectives complement one another, suggesting that the historian will do well to reflect that a modern philosopher engaging, say, with the *Essay Concerning Human Understanding*, might be responding to Locke’s ideas more or less as Locke would expect one of his own philosopher-friends to engage with it. … And the political philosopher, for his part, will do well not to underestimate the scale and density of the obstacles that stand in the way of representing the thinking of one century—particularly the engaged political thinking of one century—in the categories of another.

That one contemporary theorist enacts a methodology that I endorse is not by any means a sufficient defense of it. However, the quality of Waldron’s work is presumptive evidence in favor of the methodology, for which reason it bears mentioning. The bottom line is that I engage

---

with historical texts for the simple reason that the analytical resources they contain can be of argumentative and persuasive use for us, here and now. This is, after all, the only context in which we will have the opportunity to use them.

The other distinctive kind of text which appears throughout this dissertation are opinions from U.S. courts, principally the Supreme Court. These are often accompanied by legal scholarship, insofar as commentators on the law help illuminate the normative issues underlying the straightforwardly legal ones, and can help disentangle statutory or constitutional questions from philosophical ones. I discuss a substantial number of cases, at times going into detail about the specific conflicts that brought the parties to court. There are a couple reasons that example cases play such a prominent role. The first is that they serve to put flesh on the abstract issues we are discussing, and this may help the reader make sense of what the abstract discussion really means. Relatedly, examining these cases draws on one’s intuitions in a way that, I anticipate, will make clear that many of these cases, if not most of them, are genuinely hard cases, with meaningful practical stakes on each side. They show why and how religious exemptions matter. Finally, engaging with case law tethers our normative analysis to a particular context that it can then inform and critique. This is, after all, the aim of the project: to make some contribution to public understanding about pluralism and accommodation in liberal contexts. Showing that the normative framework can address actual cases in some meaningful way helps to demonstrate the sensibility and relevance of the framework itself.

0.6 Plan of the Dissertation

As should be clear, this dissertation is structured in the manner of one extended argument: each component contributes something to developing the stability-based approach to
religious accommodation. The first chapter sets the stage by examining and critiquing the prevailing approach to exemptions. The second, third, and fourth chapters develop the basic constructive argument, and the fifth chapter applies the framework to the specific issue of exemptions. They are not intended to stand alone, nor are they each a variation on some single theme. That said, each chapter is mostly comprehensible on its own, with the main arguments of each being more or less self-contained.

In the first chapter, I explore what I call the egalitarian approach to pluralism, identifying accounts of liberal impartiality that take equality as the guiding aim. The authors I emphasize include Adam Smith, John Stuart Mill, as well as John Rawls and contemporary exponents of deliberative democracy. For these authors (at least on some interpretations), the guiding principle for determining the scope of liberal impartiality is the equal status of citizens. I show how this account of impartiality is applied to religious exemptions, and then offer a critique of this approach. The problem, in short, is that exemptions both remedy inequalities of (substantive) opportunity and create inequalities of rights. Equality is ultimately ambivalent on the question of religious accommodation, and, I argue, cannot provide a satisfactory resolution to the issue.

In the second chapter, I identify and defend the conception of stability on which I rely as a normative ideal: a reconstructed account of Rawls’s notion of “stability for the right reasons,” which is a stability rooted in citizens’ reflective endorsement of the regime under which they live. In that chapter, I build on recent exegetical scholarship on Rawls to specify clearly what stability for the right reasons is, and then offer a defense of that normative ideal against critiques from the vantage of a “realist” movement in contemporary political theory. In short, I show that stability for the right reasons describes a type of stability that is, at least in principle if not also in fact, achievable even amid conditions of robust pluralism and disagreement.
In the third chapter, I advance a Hobbesian argument affirming that universal endorsement of a regime requires that it reflect a general norm of impartiality toward religious difference, affording persons uniform treatment irrespective of their religious beliefs. This argument depends on Hobbes’s insights regarding humans’ desire for glory and honor as a source of conflict. A regime that displays partiality among its citizens is bound to give insult to other citizens, which will both undermine those latter citizens’ affection for the regime and will stoke conflict between citizens. This argument also serves an exegetical function, contributing to literature that challenges interpretations of Hobbes as solely concerned with material interest and rational calculation. Hobbes was also acutely aware of the destabilizing potential of passion and desire for status, and his insistence on impartiality in the governance of a civil commonwealth reflects an effort to mitigate these potentials.

Then, in the fourth chapter, I identify a problem facing a regime of liberal impartiality. Namely, highly abstract and generalized principles of political organization struggle to gain enthusiastic allegiance from those who live under them. As one scholar describes it, this raises the problem of “ineffective endorsement”: even citizens who rationally affirm a political order may find their support for the regime overridden by considerations that motivate them more profoundly. In short, impartiality faces a motivational deficit. In the fourth chapter, I consider a few approaches to resolving this deficit. The first is Martha Nussbaum’s appeal to emotion as a unifying and stabilizing force. She calls for a program to cultivate citizens’ emotional attachments to the political order. A second approach is found in the role played by ethno-national solidarity and civil religion in Rousseau’s On the Social Contract. Rousseau suggests that these deep sources of identity can unify citizens in supporting a civil order that is highly general and impersonal. I identify significant weaknesses in each of these approaches, but then
present an alternative. Drawing on Locke’s *Letter Concerning Toleration* and Rawls’s notion of the overlapping consensus, I argue that support for liberal impartiality can also be grounded in citizens’ diverse beliefs, including religious beliefs and other “comprehensive doctrines.” Locke’s argument for toleration draws heavily on Christian theological claims, which suggests that citizens need not abandon their particular commitments and perspectives to affirm a common regime. On this basis, I argue that pluralism, while in some respects a challenge for stability, can also provide resources for stability.

In the fifth, and final, chapter, I return to the problem of religious exemptions. On the basis of the account of liberal impartiality from the end of stability, I formulate a framework for treating religious exemptions that is conducive to stability for the right reasons. In general, stability requires a norm of impartiality, establishing a presumption against differential treatment on the basis of religious belief. However, there are two ways in which stability may actually support exemptions for religious practice. The first is in order to avoid subjecting citizens to burdens that they cannot regard as justified. When laws and institutions infringe upon citizens’ religious practices without a good reason, those individuals feel disadvantaged. Selective exemptions can ease this subjective sense of disadvantage and shore up those citizens’ support for the regime. Second, insofar as citizens’ diverse beliefs and commitments can serve as resources to overcome the problem of ineffective endorsement, exemptions targeted at supporting the flourishing of diverse moral communities can indirectly serve stability. I engage with legal scholarship and American jurisprudence on religious accommodation to offer specific parameters for how to adjudicate exemption requests in service of these ends.
In the Conclusion, I offer a brief summary of the argument, address some of its limitations, and offer some suggestive comments on trajectories for further developing the argument from stability.

0.7 What is Impartiality?

This dispenses with the necessary preliminaries, but before proceeding to Chapter 1, we must do some work to specify at some level of precision the meaning of impartiality. Surprisingly few authors have attempted serious analyses of impartiality as a concept. Its meaning is often assumed to be obvious, and moral and political theorists typically leap immediately to the question of whether and in what contexts impartiality is required or permitted. Impartiality is assumed to mean something along the lines of suppressing “the crudest instances of self-dealing and factional oppression,” or “naked preferences.” The American Bar Association’s 2003 model code describes impartiality as the “absence of basis or prejudice in favor of, or against, particular parties or classes of parties.” The paradigm case for impartiality is the sports referee, who regulates gameplay without allowing any personal biases to influence his conduct in such a way as to benefit one team over the other. These familiar conceptualizations of impartiality focus on two elements: consistent treatment and the exclusion

---

of personal attachment. When we say that a person acts impartially, we often assume that this suffices to describe her behavior.

This imprecise colloquial understanding of impartiality is inadequate for our purposes. The first reason is that it tends to assume the moral desirability of impartiality. In actuality, impartiality is not itself a moral concept. Impartiality describes a manner of judging, deciding, and acting; the moral status of that way of acting is a separate, secondary matter. It must not be assumed that acting impartially is morally or politically appropriate. The requirement or permissibility of impartiality must be established by argument independent from the concept of impartiality itself.

Second, defining impartiality vaguely in terms of consistent treatment and the exclusion of personal preference leaves room for significant ambiguity. Consider even the apparently straightforward case of the baseball umpire. An impartial umpire does not tighten or relax the strike zone in order to advantage the team he prefers. But imagine that an umpire prefers low-scoring games, so calls a wide strike zone for both teams. Is this a violation of impartiality? It would seem not, as he treats both teams consistently. But his behavior does reflect the influence of his personal preferences. So exclusion of personal preference is neither necessary nor sufficient for impartiality. Neither can impartiality mean simply consistency, though, as an employer who consistently hired women over equally-qualified men would not be said to be impartial.

47 Gert makes this point clearly. Morality, 135–36.
48 After all, as John Kane notes, “If it is possible to do injustice through partiality, it is no less possible to do injustice with strict impartiality.” John Kane, “Justice, Impartiality, and Equality: Why the Concept of Justice Does Not Presume Equality,” Political Theory 24 (1996): 384.
49 See Gert, Morality, 133. I assume that the umpire’s strike zone is not so wide as to violate the written rules of the game. On the indeterminacy of strike zones, see Henry S. Richardson, “Our Call: The Constitutive Importance of the People’s Judgment,” Journal of Moral Philosophy 5 (2008): 17–18.
50 Gert, Morality, 133–34.
The inadequacy of colloquial understanding of impartiality becomes even clearer when considering impartiality in a political context, where the relevant actors are not only individual persons, but also institutions and impersonal laws. What does it mean, for instance, for a law to be impartial, or for an institution to act impartially? Is a zoning ordinance that prioritizes high-density growth over suburban sprawl an impartial policy, even though it reflects a clear preference for some kinds of development – and therefore some persons’ projects – over others? Or when a court must confront competing priorities (such as privacy and security, for instance) and defers to one over the other, is this an impartial decision? The familiar understanding of impartiality as consistency and/or the exclusion of personal preference cannot provide a satisfactory answer to these questions.

One prominent exception to the too-frequent casual treatment of impartiality is supplied by Bernard Gert, who sifts through a number of ambiguities troubling common understandings of impartiality and proposes a more precise definition. That definition reads as follows:

“A is impartial in respect R with regard to group G if and only if A's actions in respect R are not influenced by which member(s) of G benefit or are harmed by these actions.”

This definition offers some promise of nailing down what impartiality means. Most notably, it clarifies that it is inadequate to simply say that some person “is impartial” or “acts impartially.” Impartiality is always impartiality with regard to some specified group, and it is always impartiality in some respect or another. For instance, in a 2002 ruling on the impartiality required of judges, the Supreme Court defined impartiality as excluding bias or preference

51 Ibid., 132.
toward parties, but not necessarily including neutrality toward positions on particular issues. A judge, on this understanding, is appropriately impartial when she neither favors nor disfavors any of the parties before her; her impartiality is not undermined if she prefers some particular position on the issues at hand and allows that preference to influence her ruling. To say that a judge is impartial is to say that she exhibits impartiality toward the parties before her, in her capacity as a judge. Gert’s formulation also specifies that impartiality is a matter concerning the distribution of benefits and harms. An impartial person does not allow her actions to be influenced by how the harms and benefits resulting from those actions are distributed.

Gert’s definition has certain limitations however. First, it includes the ambiguous clause “in respect R,” for which he nowhere provides a clear definition. While the phrase might be interpreted in various ways, it seems most likely to refer to the decision context in which impartiality applies. It indicates that an actor may be impartial with regard to some group in a particular context while being partial with regard to the same group in some other context. Gert notes that a teacher may be impartial toward her students when grading their papers, but she may not be impartial when allocating her time and effort in helping them to write better papers. While I believe that this is what Gert means by “in respect R,” for the sake of clarity, I will here substitute the phrase “in decision context C.”

Second, in positing that impartiality entails indifference to “which member(s) of G benefit” from an action, Gert’s formulation fails to specify whether this indifference must apply to individual persons or to classes of persons sharing some attribute in common. It would seem that impartiality must always preclude intentionally directing benefits or harms toward specific individuals. I do not act impartially when I single out A for benefit and B for harm. But in many

---

53 Gert, Morality, 131.
cases impartiality is consistent with deliberately targeting certain classes of persons for benefit or harm. For instance, many jurisdictions impose restrictions on entertainment in the nude with the intent to prevent strip clubs from operating – thus harming those who do or would operate strip clubs. Or, the federal Child Tax Credit directs tax relief (i.e. money) at taxpayers who are raising children. In some cases, lawmakers may formulate such policies with specific individuals in mind (e.g. a city council incentivizing a company to open a facility in their jurisdiction), but this is not always the case. The federal tax code, for instance, reflects partiality toward many kinds of persons (givers to charitable organizations, capital investors, homeowners, etc.), but not toward any specific person(s).

Such laws are impartial with respect to which specific individuals benefit from them (assuming no overt corruption in the lawmaking process), but they reflect clear intentions to orchestrate the distribution of benefits and harms among classes of entities. Religious exemption rules are no different: they provide a certain benefit to the class of persons possessing religious objections to generally-applicable laws. They are impartial with regard to members of G qua specific persons, but partial with regard those same members of G qua bearers of certain attributes. This latter partiality necessarily obtains for virtually any non-trivial rule that distributes harms and benefits.\textsuperscript{54} It must specify the attribute necessary for an individual to receive either of those outcomes, thereby entailing a partiality toward the class of individuals possessing that attribute. Brad Hooker provides the example of a rule designed to reward workers for productivity: it requires providing greater benefit to those workers who are more productive. By declining to reward workers for, say, the cheerfulness with which they conduct their business, the rule remains impartial with respect to that attribute, while reflecting a

\textsuperscript{54} A trivial rule would be one designating all persons or no persons for benefits/harms. The two non-trivial rules that could avoid partiality toward any groups are a purely random distributive rule and a perfectly egalitarian one.
partiality toward those who are productive. The question always is: which attributes of persons are permitted to influence an agent’s actions with regard to those persons?

Gert’s definition of impartiality, formulated as a preclusion of influence by facts regarding “which member(s)” of the focal group benefit or are harmed by an agent’s acts, does not account for the difference between targeting specific individuals versus classes of individuals. We should thus wish to modify Gert’s formulation to accommodate this distinction. The major revision that is required is to specify that impartiality is always “with respect to” some attribute. An umpire calling balls and strikes is impartial with respect to which team the batter plays for, or with respect to the respective geographical locations of the teams, for example. The law is impartial with respect to citizens’ respective levels of wealth. A definition of impartiality must reference the attribute(s) with respect to which an actor is impartial.

In the below revised version of Gert’s definition, I have incorporated two additional changes. First, I have replaced Gert’s “with regard to” to “toward” in order to more clearly capture the group affected by an agent’s actions. This change also avoids confusion with “with respect to,” which I use to identify the attribute that does not influence an agent’s actions. Second, I wish to specify that this definition describes impartiality toward persons, in order to distinguish it from a further kind of impartiality, which I describe below. Thus, a revised version of Gert’s definition of impartial conduct might read:

\[ I_{\phi} \text{: A is impartial with respect to attribute } X \text{ in decision context } D \text{ toward group } G \text{ if and only if A’s actions in decision } D \text{ are not influenced at all by whether and in what amount} \]

the class of members possessing X receives the benefits or harms that result from A’s actions.

While this revision captures more precisely the relationship between impartiality and the recipients of benefit or harm, it nevertheless remains incomplete. Specifically, Gert’s formulation captures only those decision contexts where benefits and harms consequent to an agent’s actions are at issue. Many decisions do not have this feature, and impartiality is a relevant category in at least some of those cases. Imagine, for instance, that I am driving my car in an unfamiliar city, carrying two passengers: my sister and her new partner, whom I do not like. My sister suggests that I take route A to our destination, because it is quick; her partner suggests route B, because it is more scenic. I am forced to choose between them. Selecting one or the other confers no benefit or harm on either my sister or her partner, but I may nevertheless decide the matter impartially by setting aside any kinship-based preference for my sister, and instead deciding on the basis of my preferences regarding speed and beauty.\textsuperscript{57} Thus, it would seem that a definition restricting impartiality to benefitting and harming actions is too narrow. It would seem that an easy clarification is possible: the “benefit” in this case is the status of being selected, and it is conferred not on one passenger or the other but rather on one of their proposals. But this solution stretches the concept of benefit untenably. More importantly, it obscures an important feature of impartiality: it may apply to reasons as well as to persons.

Impartiality with regard to reasons is analogous to impartiality with regard to persons, insofar as it entails excluding some factors from influencing one’s decision in a particular context. Precluding the influence of a reason is a complex psychological feat, but it essentially

\textsuperscript{57} For the sake of simplicity, I assume that neither route has any decisive advantage over the other, and neither passenger is of such flimsy ego that my choice will appreciably affect their sense of self-worth.
involves setting aside a reason, ignoring it, or excluding it from consideration. Despite its difficulty, this is not an unfamiliar exercise: juries do it when they are instructed to ignore some out-of-order statement by a witness. We can model the exclusion of a reason by saying that an actor excludes a reason when the complete reason for her action does not depend on the truth of that reason. In the case of my two passengers offering conflicting directions, I am impartial with respect to the fact that I know and trust my sister (which is a reason for taking route A) insofar as my decision would be the same even if it were not true that I know and trust my sister. Impartiality among reasons must always be in regard to some subset of all the reasons pertaining to an agent in some decision context. To act with impartiality toward all one’s reasons would be necessarily to act randomly or irrationally. We can define impartiality toward reasons as follows:

\[ I_R: A \text{ is impartial with regard to a subset } G \text{ of her reasons for action in decision context } D \]

if and only if A’s actions in decision D are not influenced at all by which member(s) of G are true or false.

Captured in this way, it becomes clear that impartiality with regard to reasons is a more abstract formulation of impartiality of which impartiality toward persons (captured in \( I_P \)) is a specific instance. To act impartially among persons is to exclude from the reckoning of one’s balance of reasons facts concerning which persons will be benefitted or harmed by my actions. That person P possesses some attribute X can be a reason for A’s action. Impartial conduct involves setting aside reasons of this sort. Impartiality as the exclusion of reasons is reflected in the following gloss offered by Joseph Raz:

---

People are impartial if and only if, in matters affecting others as well as possibly themselves, they act on relevant reasons and shun irrelevant ones, in particular … irrelevant considerations that favor themselves or people or causes dear to their hearts, and if their evaluation of the situation is not distorted by the fact that such people or causes are dear to them.59

Typically, as Raz suggests, impartiality involves setting aside personal attachments when deliberating over action. But not necessarily: impartiality is a more formal notion describing the exclusion of any reason(s) whatsoever. Raz’s definition suggests, though, that impartiality is concerned with distinguishing between those reasons that are “relevant” and those that are not. The relevance of certain reasons is a function of the decision context, and in many cases context alone will indicate clearly whether a given reason qualifies. A batter’s ethnicity is, for instance, clearly irrelevant to an umpire’s determination of ball or strike. The rules of the baseball supply the parameters of relevance, and all reasons for an umpire’s action that do not derive from the rules are to be “shunned.” In other cases, however, relevance will be difficult to judge. Is a defendant’s material poverty relevant to his criminal punishment? The key to impartiality, therefore, is determining which reasons are relevant and which are irrelevant in a given decision context. In terms of IR, this means specifying the subset of reasons whose truth value is to be disregarded in reasoning. For impartial person-affecting conduct, it is those distinctions between (classes of) persons that are judged to be irrelevant that are properly precluded from influencing the agent’s action.

At this point it might seem parsimonious to drop the person-affecting definition of impartiality (I\textsubscript{P}), as it is essentially a specification of the more general definition of impartiality as discrimination among reasons (I\textsubscript{R}). For two reasons, however, it will prove useful to retain the two formulations. The first reason is that impartiality is most frequently invoked in and is typically understood to pertain to decision contexts involving person-affecting conduct. The baseball umpire, the judicious parent, the fair-minded employer – all of these are in the business of distributing benefits and harms. I\textsubscript{P} helpfully characterizes what it means to act impartially in such contexts.

The second reason to retain the two formulations of impartiality is that they correspond neatly to the two principles that, I will argue, characterize liberal impartiality in political theory. One way to identify these principles is to consider three respects in which impartiality may function in relation to rules generally: (1) application of a rule, (2) a rule’s content, and (3) justification of a rule. I leave open for now the question of whether impartiality in any of these three exercises is ever permissible or required.\textsuperscript{60} What is important is that in performing each task, impartiality is possible.

The first two – application and content – are clearly person-affecting enterprises: they determine how harms and benefits are distributed among persons under the rule’s jurisdiction. The third – justification – is a matter of reasoning: it concerns the grounds that provide support for the rule. I will leave impartiality in the application of a rule aside here, as it is not especially interesting from a theoretic perspective. For rules the content of which are impartial (with respect to X), the ordinary application of that rule ought to be impartial (with respect to X). To fail in applying an impartial rule impartially is to fail in applying the rule. (Though this kind of “failure” may be justifiable in some cases.) For rules that are not themselves impartial, it is not

\textsuperscript{60} See, e.g., Hooker, “When Is Impartiality Morally Appropriate?”
clear what significance impartiality of application would have. The impartial application of rules is of course a challenge and a priority in the real world, but it is not theoretically interesting. For our purposes, we may identify the two provinces of impartiality as those of distribution (of harms and benefits) and justification (of distributions). ⁶¹

Here we are focused on a specific political species of rule: laws. I assume that a political order consists of a system of rules whose function is to distribute harms and benefits – whether these be material benefits, criminal sanctions, or powers and rights. ⁶² What does it mean for a law to be impartial in its distribution of harms and benefits? After all, a rule does not “act” as an agent does, but a rule does prescribe a distribution of benefits and harms. Impartiality for a law, with respect to some attribute, then, must consist in a distribution that does not target persons possessing that attribute for special benefit or harm. We might capture this feature as follows:

Law L is distributively impartial toward the members of some group G with respect to attribute X if and only if it distributes harms and benefits without respect to the distribution of attribute X among the members of group G. ⁶³

Distributive impartiality is essentially a matter of distributing harms and benefits in a manner that does not depend on whether persons possess some given attribute. We might say that

the fact of their possessing X is excluded from consideration, that it is “shunned,” or that it is

---


⁶³ We could specify this further by saying that “there is no person P for whom P’s possessing X implies any consequence for the distribution of harms and benefits prescribed by L.”
hidden behind a veil of ignorance. The point is that the distributive prescription does not take into account the presence or non-presence of attribute X in those whom it governs. The distributive impartiality of a law tracks definition of impartiality toward persons, where the distribution of harms and benefits eschews preferences for certain persons or classes of persons.

Justification involves the reasons that support the distributive scheme prescribed by a law. Typically in contemporary liberal theorizing, the justification at issue is *public* justification, or the justification of a law to the persons who are subject to it. In general, a law or principle is justified to someone when that person has a sufficient reason to endorse it. This gloss leaves things seriously under-specified, but will suffice for now. Justification may involve any number of reasons, and also different *kinds* of reasons – prudential, moral, scientific, religious, etc. Impartial justification is a matter of indifference toward the truth or falsity of some set of reasons. A law is impartial with respect to religious commitments, for instance, when no particular religious premise is required in order for the law to be justified. Impartiality in the justification of a rule tracks impartiality with respect to reasons, in the sense that it involves excluding certain reasons from influencing one’s actions, or adopting a posture of neutrality with regard to the truth value of those reasons. Specifically,

Law L is justificatorily impartial with respect to set of reasons S if and only if there is no member M of S for which the fact that M is false implies that L is unjustified.

Having now formulated definitions of what it means for a law to be impartial in its distributive prescriptions and in its justification, it remains only to link these with the normative

---

64 I here follow Kevin Vallier: “A coercive law L is justified only if each member I of the public P has some sufficient reason(s) Rᵢ to endorse L.” *Liberal Politics and Public Faith: Beyond Separation* (New York: Routledge, 2014), 24.
commitments to impartiality that are characteristic of liberal theory. My intention here is not to defend these commitments, but rather merely to identify them in order to give content to what I take to be the core normative theses of liberal impartiality. To be sure, not all liberals endorse impartiality, nor do all liberals endorse impartiality precisely as I intend to articulate it here. Nevertheless, impartiality is a pervasive and central theme in liberal theory, and the principles I shall momentarily supply are sufficiently content-indeterminate to accommodate a wide range of formulations. In Chapter 1, I will show how these elements of impartiality have figured prominently in liberal thought, and will outline a justificatory basis for them grounded in the equal status of citizens. In Chapter 3, I will advance an alternative defense of impartiality on the basis of the end of stability. For the sake of linguistic simplicity, I have streamlined the language in the two principles, which read as follows:

*Distributive Impartiality*: Political arrangements ought to distribute benefits and burdens impartially with respect to politically irrelevant differences among persons.

*Justificatory Impartiality*: Political arrangements ought to be supported by a justification that is impartial with regard to comprehensive doctrines.

For Distributive Impartiality, I have shifted focus from individual attributes to differences between persons. Logically, these are equivalent (persons differ in that some persons possess X and others do not possess X), but given the present focus on pluralism, we are more helpfully served by a frame that pinpoints the issue as the ways that persons *differ*. The phrase “politically irrelevant” is deliberately unspecified, leaving room for various accounts of which differences
Among persons are of political significance sufficient that they ought to factor into the distribution of benefits and harms. For Justificatory Impartiality, I have specified that the relevant set of reasons in respect to which justification should be impartial are those deriving from comprehensive doctrines. There may be other sets or kinds of reasons that could reasonably be specified here, but the predominant focus in theories of public justification has been on those reasons that flow from citizens’ metaphysical, moral, or religious beliefs. In keeping with the generic treatment of justification noted above, I leave open questions regarding who is presumed to be constructing or offering a justificatory account, as well as to whom it is presumed to be offered.

A final note is in order before concluding this stage-setting regarding the concept of impartiality. In this project I am focused specifically on political impartiality, which I take to consist in the principles of distributive and justificatory impartiality. I do not take up the question of moral impartiality – that is, to what degree morality requires impartiality in judgment, affection, or conduct, or to what degree impartiality is morally permissible. The well-known test case for moral impartiality involves a person who finds two children drowning: her child and a stranger’s child. Assuming she can save only one child, is she permitted to prefer her own? Most people will assume that some form of impartiality is a necessary ingredient of morality, though Bernard Williams has famously derided the notion of saving a stranger’s child over one’s own as “one thought too many.” Whereas questions over impartiality in morality may strike some as unnecessary handwringing, it seems indisputable that impartiality has an important role in

---

65 Rawls defines a comprehensive doctrine as one that “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.” *Political Liberalism*, 13.

politics. The specifics of that role are, in contrast, the subject of much dispute. It is into that
disputed space that this dissertation aims to enter.
CHAPTER ONE:

THE EгалITARIAN BASIS OF IMPARTIALITY

The goal of this chapter is to identify and explore what I have called the predominant approach to impartiality as a response to pluralism in liberal theory, namely the egalitarian approach. By and large, the norm of impartiality has been defended on the basis of a commitment to respect persons as equals. Of course, accounts differ on exactly what that commitment consists in, and precisely how impartiality is supportive of or required by that commitment. Also, equality need not be understood to provide grounding for only impartiality: one could argue, for instance, that in order to respect equality, political institutions ought also to be democratic. Here, though, I am concerned to show that equality can serve – and frequently has served in liberal theory – as a foundation for the political norm of impartiality.¹

This chapter covers a good deal of ground, much of it only superficially. The reason for this is that I am less interested here in investigating any specific invocation of equality than in tracing the prevalence of viewing equality as a foundation for impartiality. To do this, I first highlight three key appearances of impartiality in liberal theory: first in Adam Smith’s theory of moral sentiments (including his famous figure of the impartial spectator), then in J.S. Mill’s utilitarian theory, and finally in neo-Kantian contemporary liberal theories of justice, of which Rawls’s justice as fairness is the primary example. Each of these, I claim, supports a prescriptive conclusion of impartiality of one sort or another. Of course, there are other examples that might have served equally well. But Smith, Mill, and Rawls represent pivotal figures in the liberal

¹ To say that equality serves as a “foundation” or a “ground” for impartiality is not to say that equality alone motivates the general principles of liberal impartiality, as if impartiality is a necessary conceptual entailment of a normative commitment to equality (however understood). Rather, equality serves as a guiding end that, in conjunction with various other independent premises, supports an argument that concludes in impartiality.
tradition, and they function here to demonstrate the centrality of the egalitarian basis of impartiality in liberal theory. Of course, each thinker here intends or understands impartiality in slightly different terms, and I make no claim that each version of impartiality is identical with the others. Nor, as will become clear, do I suppose that each thinker grounds impartiality in equality in an identical manner. Rather, I merely wish to show how some form of impartiality figures in these key moments in the liberal tradition, and how at each turn it is grounded in some version of a commitment to equality.

After surveying these appearances of impartiality in liberal thought, I shift to a more abstract mode and formulate what I call a generic egalitarian argument for impartiality. I offer this construction in order to demonstrate clearly how the premise of equality serves to underwrite the normative commitment to impartiality – including both distributive and justificatory impartiality. Establishing clearly how equality motivates impartiality in this chapter will establish a point of contrast with the stability-based argument I will advance in chapter 3. To conclude the present chapter, I offer a weak critique of the egalitarian basis. It is a “weak” critique not because (I believe) it is faulty or invalid, but rather because it does not pretend to entirely debunk or disprove the egalitarian account. Instead, I aim to show that the egalitarian understanding of impartiality is not up to the task of resolving the challenge of religious exemptions that I identified in the introduction. Specifically, equality generates contradictory conclusions, grounding both support for and opposition to religious exemptions. It is, in this sense, profoundly indeterminate, and this indeterminacy lends credence to my suggestion that an alternative foundation for impartiality – and a corresponding framework for treating religious exemptions – merits exploration. This conclusion sets the stage for introducing the end of stability and, ultimately, advancing an argument for impartiality on the basis of that end.
Before proceeding to consider the egalitarian basis for impartiality in Smith, Mill, and contemporary liberals, recall the definition of impartiality offered by Joseph Raz. For Raz, impartiality consists in appropriately discriminating among those reasons which merit consideration and those which do not. The latter considerations are to be “shunned.” The critical question, then, is: which reasons are to be shunned? In other words, with respect to what ought one to be impartial? Each of the accounts below supports impartiality in some slightly different respect, but each insists on the shunning of some particular consideration(s) in moral judgment, the formation of political policies, or public justification of political arrangements. Giving content to any account of impartiality requires specifying which considerations are to be accommodated and which shunned – and Mill, Smith, and Rawls each do this in their own way. The ultimate problem, which I will highlight toward the end of the chapter, is that the premise of equality itself provides no clear indication how to draw these boundaries, which leaves it susceptible to various, often contradictory, formulations.

1.1 Moral Sentiments: Smith’s Impartial Spectator

One prominent instance of impartiality in liberal moral and political thought is found in Adam Smith’s famous figure of the “impartial spectator,” elaborated in his second most famous work, The Theory of Moral Sentiments. For Smith, impartiality is captured in the perspective that a person (even an imaginary one) occupies by virtue of a particular kind of situation with respect to the actors in a given interaction. Moral judgment is properly executed by assuming a vantage of impartiality vis-à-vis the object of one’s judgment.

---

Smith’s moral theory takes off, actually, from the idea of sympathy, rather than equality or impartiality, per se. Sympathy has to do with a person’s feeling of concern for the well-being of another – what Smith calls “fellow-feeling.” When a person takes note of another’s feelings, she will often “feel something which, though weaker in degree, is not altogether unlike them.” This occurs not through some immediate sharing, but rather by a function of the imagination. When we observe another person who is caught up in some particular emotion or sensation, we tend to imagine ourselves in his or her situation, and thereby acquire a sense of what the other is feeling. Smith affirms that it is impossible to access the sensations of another directly, but by means of imagination one can “enter as it were into [another’s] body, and become in some measure the same person with him.” In ordinary use, sympathy typically refers to sharing in sorrowful emotions, but Smith intends it to describe sharing in any of another’s sensations – “any passion whatever.” That said, he does not imagine that sympathy tends naturally to arise in equal portion in response to all passions. Moreover, sympathy does not mean feeling precisely the same emotion as another person does. This is because, Smith holds, sympathy “does not arise

---


7 Ibid., 15.

8 Ibid., 15–16, 56.
so much from the view of the passion, as from that of the situation which excites it.”

Sympathy is not an act of mimicry, but of imagination.

Sympathy itself does not comprise any evaluation of the agent’s actions in her situation, but it does provide the basis for judgment in Smith’s system. To wit: if the spectator, upon imagining herself in the place of the agent, feels similarly to the agent, the spectator is said to approve of the agent’s passions. If the spectator’s response to the situation differs from that of the agent, the spectator disapproves. In short, “To approve of the passions of another, therefore, as suitable to their objects is the same thing as to observe that we entirely sympathize with them; and not to approve of them as such, is the same thing to as to observe that we do not entirely sympathize with them.” Approval and disapproval, in Smith’s system, represent consonance or dissonance in the passions of the agent and spectator, respectively. This comparison is the basis for evaluating the actions of another: when an agent takes an action motivated by passions that the spectator, sympathizing with the agent, does not feel, the spectator judges this action improper. Conversely, when the agent acts on passions also felt by the spectator, the spectator judges the action to be proper.

Of course, the coincidence of passions in actual persons is an obviously inadequate basis for moral evaluation. That I judge another’s passions improper (because they do not match what I imagine my own passions would be in his situation) very possibly reflects my inability to perfectly sympathize with his situation, rather than any objective impropriety in his passions. Actual spectators are burdened by interests, prejudices, and preferences that shape the passions they imagine feeling in another’s shoes. The fatal problem with the evaluations of actual spectators is that they tend to be shot through with partiality. In Smith’s words, “So partial are

9 Ibid., 16.
10 Ibid., 22.
11 Ibid., 24–25.
the views of mankind with regard to the propriety of their own conduct, both at the time of action and after it; and so difficult is it for them to view it in the light in which any indifferent spectator would consider it.” The failure of actual spectators is their inability to judge impartially.

Hence, the figure of the impartial spectator. The appropriate way to examine one’s own conduct, for Smith, is to imagine how it would be assessed by an observer who held no stake in the matter. As with the basic phenomenon of sympathy, this requires an act of creative imagination: an agent imagines herself in the position not of an actual other, but of a hypothetical impartial observer, and considers whether this spectator would approve or disapprove of the agent’s conduct. The impartial spectator thus personifies a perspective free of bias and self-interest. Smith writes, “it is only by consulting this judge within, that we can ever see what relates to ourselves in its proper shape and dimensions; or that we can ever make any proper comparison between our own interests and those of other people.” The impartial spectator offers a morally trustworthy vantage, and for Smith, this is the proper basis of moral judgment.

Why, though, is the perspective of this impartial observer privileged over those of actual, interested persons? Sure, the impartial spectator’s perspective is distinctive, but Smith’s characterization suggests that it is also hierarchically superior to other moral vantages. What is the reason for preferring impartiality, on Smith’s view? Very simply, Smith holds that impartiality, modeled by the perspective of a disinterested spectator, is a means of regarding persons as equals. He writes that the impartial spectator “calls to us … that we are but one of the multitude, in no respect better than any other in it; and that when we prefer ourselves so shamefully and so blindly to others, we become the proper objects of resentment, abhorrence,

---

12 Ibid., 182.
13 Ibid., 156–57.
Humans tend naturally to prefer themselves – to prioritize their own concerns and to discount the interests of others. Partiality in moral judgment is flawed on account of its inegalitarian – namely, self-preferring – influence. At the same time, the offense we take at others’ conduct derives primarily from their own self-preference:

What chiefly enrages us against the man who injures or insults us, is the little account which he seems to make of us, the unreasonable preference which he gives to himself above us, and that absurd self-love, by which he seems to imagine, that other people may be sacrificed at any time, to his conveniency or his humour.15

The sympathetic process is, for Smith, the proper grounding of morality, but it is only the (imagined) sympathetic response of a (hypothetical) impartial spectator that is morally trustworthy. The function of the impartial spectator is to remind individuals of the basic equality of persons, such that they might adjust their moral judgment accordingly. Impartiality instantiates equal regard in moral judgment. Some interpreters have even read the centrality of equality in Smith as evidence of his thoroughgoing Kantianism. For example, Samuel Fleischacker draws explicitly the connection between Smith and Kant on this point, writing that “For Smith as for Kant, then, regarding all human beings as equal is essential to morality.”16

Moreover, Smith’s focus on impartiality as a reflection of equality is not limited to Theory of Moral Sentiments – it figures importantly as well in The Wealth of Nations. Though that text is

14 Ibid., 159.
15 Ibid., 96.
ostensibly concerned with free individuals pursuing their respective interests, Stephen Darwall writes that “the theme of equal dignity is in the background.”

The notion of an impartial spectator has been influential in moral theory insofar as it prefigured and inspired doctrines such as Roderick Firth’s conception of the “ideal observer” or Thomas Nagel’s “view from nowhere.” These views take the view that moral judgment is properly exercised from a position of disinterest, for the fundamental reason that disinterest conduces to egalitarian reasoning. To be sure, these are primarily moral, rather than political doctrines, and their political implications are not obvious. So, there is a danger in conflating moral and political impartiality. But from Smith’s moral account of the impartial spectator, one can infer at least a presumption in favor of impartiality in politics.

1.2 Utilitarianism: Equal Regard for Interest

Smith’s *Theory of Moral Sentiments*, as the title would suggest, offers a moral theory, rather than an overtly political one. The egalitarian basis for impartiality is on display in a context more proximate to the political in the utilitarianism of John Stuart Mill. The motivation driving the development of utilitarian ethics by such figures as Jeremy Bentham and James Mill was to orient morality around the consequences produced by actions, rather than their conformity with some natural order or set of moral rules. Moreover, utilitarianism adopted a hedonic measure of good, whereby it is the pleasure produced (or reduced) by an action that constitutes its moral worth. Whereas Bentham held that all pleasures are of equal merit, J. S. Mill relied on a

---

distinction between higher and lower pleasures in order to avoid the conclusion that push-pin is as good as poetry, if the two activities produce the same quantities of pleasure.

Impartiality plays an important role in utilitarian moral theory. A basic commitment of utilitarianism is that calculations of utility are to proceed without preference for or bias toward some individuals over others. That is, the pleasures of one person must be regarded as carrying equal weight as the pleasures of another. This requirement of equal consideration includes an agent’s weighing of her own pleasure against that of some other person. Mill writes “As between his own happiness and that of others, utilitarianism requires him to be as strictly impartial as a disinterested and benevolent spectator.” Notably, Mill invokes the perspective of an ideal impartial observer, though in contrast to Smith, Mill introduces the condition that the spectator be not only impartial but also benevolent (presumably to motivate, in addition to the spectator’s ability to judge utility accurately, his interest in increasing aggregate utility). For Mill, impartiality is not sufficient for morality, but it is decidedly necessary. Utilitarianism requires impartiality at the foundational level of regard for interest.

The insistence on impartiality in utilitarian moral reasoning reflects an implicit egalitarianism. No person’s pleasure is more valuable, important, or worthy of pursuit than that of another. Just as Smith’s impartial spectator served to model equal regard, calculation of increases or diminutions of aggregate utility puts persons an equal plane, with the utility of each person being neither privileged over nor subordinated to the utility of any other. The fundamental role of equality in grounding impartiality toward agents and interests that Mill calls

---

for is continued in later utilitarians, such as Henry Sidgwick. Sidgwick describes utilitarianism as grounded fundamentally in

the self-evident principle that the good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other; unless, that is, there are special grounds for believing the more good is likely to be realised in the one case than in the other. And it is evident to me that as a rational being I am bound to aim at good generally,—so far as it is attainable by my efforts,—not at a particular part of it.21

Sidgwick does note one legitimate reason for prioritizing the good of one over some other, but it is a prudential consideration: the relative likelihood of achieving various persons’ interests. But the respective interests themselves are of basically equal worth. Sidgwick continues that the principle of benevolence requires that “each one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him.” Thus, for Sidgwick as for Smith, moral judgment requires an agent to adopt a perspective of impartiality with regard to her own situation and conduct in order to formulate proper moral assessments. Failure to do this is a failure to regard one’s own interests as equal to those of others.

The political upshot of this utilitarian view, as Mill famously elaborates in On Liberty, is a program of impartiality toward preferences and ways of life. Political rule must, according to

Mill, permit wide latitude for individuals to pursue their “own good in [their] own way.” Mill’s famous harm principle requires that the instruments of the state refrain from regulation of conduct that does not impose upon other citizens. As Mill writes, “Over himself, over his own body and mind, the individual is sovereign.” That is, the political order must take a hands-off stance toward citizens’ individual pursuit of the good. Moreover, Mill links the moral impartiality of utilitarianism with the political principle that individual interests are to be given equal weight in determination of the laws and arrangements that bear on those people. He writes that “Society between equals can only exist on the understanding that the interests of all are to be regarded equally.” Of course, the equal consideration required by utilitarianism at the level of weighing interests does not necessarily entail equal treatment under political institutions or strict equality along metrics of wealth or well-being. The problems of “utility monsters,” redistributive organ transplants, or, by extension, oppressive social practices that manage to nudge the overall utility metric slightly upward have been well-noted. So equal consideration of interest need not lead to equality of treatment or equality of condition under political institutions. Though, Mill did notably advocate advances in the social and political standing of women, reflecting a degree of indifference to the quality of sex as a determinant of fitness for public life. My point, however, is not to comment on Mill’s personal commitment to political equality. (Which his endorsement of weighted voting in Considerations on Representative Government

---

23 Ibid., 14.
might appear to contradict. Rather, I mean to highlight the basic role of equality in motivating the impartiality of Mill’s moral and political utilitarianism. For Mill, it is equal regard for the interests of each individual that grounds utilitarianism’s impartial consideration of utility.

1.3 Liberal Rationalism: Rawls’s Original Position and the Idea of Public Reason

In the twentieth century, arguably the most notable defense of impartiality in liberal political theory was made by John Rawls. Rawls turned pointedly away from utilitarianism, which by the time of his writing had dominated moral theory for most of the century. His major critique of utilitarianism was that utilitarianism prioritizes the aggregation of interests without due attention to the distribution of their satisfaction. As I indicated above, utilitarianism seems inevitably to endorse pernicious or at least inegalitarian arrangements so long as they increase overall utility. Rawls describes this problem as a failure to “take seriously the distinction between persons.” In short, Rawls objected to the fact that utilitarian inequalities, while justified from the perspective of the aggregate whole, would be unjustifiable to those on the losing end of the bargain. While utilitarian reasoning affirms equality by adopting an impartial stance toward the beneficiaries of utility, Rawls objected that this kind of impartiality actually runs counter to the fundamental distributive requirements of equality. Rejecting utilitarianism on these grounds, Rawls’s Theory of Justice moved in the direction of affirming not merely equality of interests, but of persons themselves. The basic assumption of Rawls’s “justice as fairness,” after all, is a conception of persons as free and equal.

---

30 Ibid., 131–32.
This basic commitment to equality is reflected in Rawls’s theory by his signature conceptual product: the device of the original position. The original position is a hypothetical bargaining situation in which parties negotiate terms of social cooperation from behind a “veil of ignorance,” which precludes knowledge of many facts regarding each individual’s particular characteristics, aptitudes, and social position. By reasoning about desirable social arrangements under such epistemic constraints, the parties are positioned to select political principles that are not biased in favor of particular social groups and that constrain inequalities. Thus, the original position models the conditions of rational impartiality that are appropriate to the deliberative cooperation of persons who stand toward one another as free and equal persons. As with Mill’s utilitarian view, Rawls’s justice as fairness does not insist on the absolute realization of equality between citizens. Rather, as specified in Rawls’s famous “difference principle,” inequalities are permitted, but only insofar as they serve to increase the absolute well-being of the least well-off persons in society. Thus, for example, inequalities deriving from free economic exchange may be permissible if free economic exchange generates overall economic growth that results in benefits to those who are relatively poorest. The major difference, then, between the inequalities permitted under justice as fairness versus those endorsed by utilitarianism is that Rawls insists that inequalities must be justifiable not merely from a perspective of aggregate utility, but from the perspective of particular persons: those on the “losing” end.

One innovation of Rawls’s theory was the prominent role it granted to justification. More than equality of opportunity, equality of outcome, or impartiality of treatment, Rawls’s theory emphasized the fact that the principles of justice specified in justice as fairness could be justified to parties negotiating an (idealized) agreement. Of course, the notion of politics by agreement is as old as the social contract tradition, but Rawls’s contribution was to carry the schema “to a
higher level of abstraction.”31 This increased abstraction – namely, conceiving of the contract negotiation as occurring in a hypothetical “original position” in which parties are severely limited in knowledge about themselves and the lives they will lead – serves to heighten the degree of impartiality that a suitable justification for a political principle must possess. A person who does not, by virtue of the veil of ignorance, know exactly what kind of person she is will accept only justifications that would apply to the full range of her life possibilities. That is, she will accept only those justifications that are impartial with regard to the particulars of her personal characteristics and life situation.

In this, we see a contrast between Mill’s attention to the generation of pleasure without regard to location of that pleasure among persons and Rawls’s focus on distribution of benefits (what he calls the “primary goods”) without regard to “morally arbitrary” personal characteristics.32 Whereas utilitarianism is impartial with regard to personal identity, justice as fairness is impartial with regard to personal difference. All of this, we must recall, stems from Rawls’s primary commitment to elaborating the political conditions appropriate for a system of social cooperation among free and equal persons. Relating to one another as free and equal citizens, for Rawls, requires that we come to agree on political principles that are impartially justified.

The foregoing discussion has focused on Rawls’s early work in A Theory of Justice. In his later work, the mechanisms of justification shift, but the fundamental role of equality remains paramount. Indeed, justification as a requirement of equality takes on an even more prominent role. In Political Liberalism, Rawls drops his emphasis on the original position and instead devotes his attention to a set of concepts and norms that can be applied to actual (as opposed to

31 Ibid., 3.
32 Ibid., 61–65.
hypothetical) political exchange. The idea of the overlapping consensus describes the phenomenon by which a conception of justice that is freestanding of any particular “comprehensive” moral or metaphysical doctrine can win agreement by others in society, each of them relating it to their own particular comprehensive views. The idea of public reason, similarly, describes the norms of discourse that structure a public sphere governed by “political” (as opposed to “comprehensive” doctrines). In other parts of this project, I will explore Rawls’s views in *Political Liberalism* as motivated by the concern of stability (stability “for the right reasons,” specifically), but it is the case that stability is not the sole end in Rawls, and traditional interpretations of Rawls taken free and equal citizenship as the primary focus, to which stability serves a complementary or ancillary function. Given this, it is worth taking a moment to consider the work done by equality in this part of Rawls’s work.

For Rawls, the idea of public reason supplies norms of public justification that are appropriate to free and equal citizens engaged in collective deliberation and decision-making. Rawls says that these norms specify the form of the “political relationship” among citizens of a liberal society. Public deliberation is principally a justificatory project that involves citizens offering one another reasons to support this or that political arrangement. The paradigm of politics on this view is highly rational, and rests on a hope that political decisions can be arrived at via a process of collective reasoning. Under Rawls’s conception of public reason, citizens are to offer one another justifying reasons for laws and political arrangements, but they are to refrain from employing their own comprehensive views in doing so. That is, citizens must ground their appeals in ways that do not depend on controversial conceptions of the good or “thick” moral or metaphysical doctrines. Rawls writes, “Public justification is not simply valid reasoning, but

---

argument addressed to others: it proceeds correctly from premises we accept and think others could reasonably accept to conclusions we think they could also reasonably accept.”

Elsewhere, Rawls writes that this requires appeal to justificatory reasons that are “accessible to citizens’ common reasons” – i.e., they require no special epistemic faculty or system of belief in order to appreciate their force. Rawls explains the significance of public reason by noting that reasonable citizens are those who regard one another as free and equal participants in a system of social cooperation, and who are willing to reason reciprocally (“publicly”) about the terms of that cooperation on the basis of “political” conceptions of justice that can be mutually endorsed. Thus, the doctrine of public reason elaborates norms of social cooperation that are appropriate to persons who are, and who regard one another as, equals.

Rawls’s theory of political liberalism paralleled similar accounts advanced by a few other thinkers (Charles Larmore, for example), and along with those has given rise to a wide body of literature expounding and examining the formulation of public reason’s aims and norms. As I have tried to show, these views rest heavily on conception of persons as free and equal. The normative commitment to equality has been my primary focus, but autonomy merits notice as well. In particular, equality and autonomy jointly motivate the imperative of public justification.

Given a claim to freedom, coercion must be justified on some grounds. Given equality, no person is by default justified in exercising coercion. The only way that coercion – including political arrangements, which are inevitably coercive – can be justified among free and equal citizens is if it is justified to the persons who are subject to that coercion. In short, the status of persons as

35 Ibid., 155.
free and equal requires that coercion meet a standard of justification to the object of that coercion.

A number of theorists have captured this feature of liberal thought under the banner of “respect,” along the lines of the Kantian notion that persons are to be treated as ends in themselves and never simply as a mere means. Larmore writes that “to respect another person as an end is to require that coercive or political principles be as justifiable to that person as they presumably are to us.”38 Conversely, to coerce someone – or, in a political setting, to endorse coercive laws – absent a justification that could reasonably be appreciated by that person is to fail a duty of respect to them. Importantly, because persons are equal, they are equally deserving of this respect, and therefore the standard of justification applies with respect to all of them. Coercion must be justifiable not only to “us” or to some objects of coercion, but to all of those who are coerced. The fact of pluralism constitutes a critical premise in this argument. In a social context where citizens cannot be expected to universally share common comprehensive doctrines (such as religious belief systems), to offer a justification rooted in any one particular doctrine is to fail to offer a reason that those who do not affirm that doctrine could endorse. That is, if I offer as justification for coercion the reason that “the Bible says so,” someone who does not share my belief in the truth of that text cannot be expected to accept my justification. If the presumption against coercion requires justification by reasons that can be shared among all citizens, then religious reasons, along with other appeals to comprehensive views, must be excluded, as they cannot achieve such wide endorsement. The issue is not merely that religious reasons are unlikely to win actual support (though this may be true); it is that religious reasons are not amenable to the kind of meaningful evaluation that would allow other citizens to engage

reciprocally with them. As Rawls has put it, they are not “accessible to citizens’ common reasons.” Respect for one’s fellow citizens requires that one eschew such appeals in the enterprise of public justification.

Thus, the ideal of public reason, in which justifying reasons are to be articulated on bases that can be widely accepted. Other authors – in addition to Rawls and Larmore – who have written in this vein about the norms of public justification that follow from a commitment to equality include Gerald Gaus, Stephen Macedo, and Fred D’Agostino, to name just a few. A related body of literature has grown up around the notion of “deliberative democracy” – a close relative of the public reason school in contemporary liberalism. For example, Jürgen Habermas has elaborated a well-known normative account of impartial democratic discourse centered on his discourse principle, which specifies that “just those action norms are valid to which all possibly affected persons could agree as participants in rational discourses.” A normative commitment to equality plays a fundamental role in motivating this central principle. Habermas writes that the principle affirms those norms “that can be justified if and only if equal consideration is given to the interests of all those who are possibly involved.” Thus, the terms of democratic deliberation are given by the basic assumption of equality. The arguable leader of the deliberative democracy school in American philosophy is Joshua Cohen, who has written that

40 Rawls, Justice as Fairness, 90.
“The notion of a deliberative democracy is rooted in the intuitive ideal of a democratic association in which the justification of the terms and conditions of association proceeds through public argument and reasoning among equal citizens.”

Together, the schools of “public reason liberalism” and deliberative democracy have dominated much of contemporary liberal theorizing. Without diving too deeply into the nuances of various authors writing along these lines, I have attempted to show how the norm of equality features as a central normative commitment of these theories of political discourse and public justification.

1.4 A Generic Argument from Equality to Impartiality

For each of the thinkers above, a commitment to equality plays an important role in their respective defenses of political impartiality. The purpose of surveying Smith, Mill, Rawls, and Habermas on this theme is to demonstrate the predominance of egalitarian accounts of impartiality in liberal theory. Any number of other thinkers, of course, might have served equally well to show the centrality of equality in theories of liberal impartiality. Brian Barry, for example, advocates a theory he calls “justice as impartiality,” in which the preferred principles of justice “capture a certain kind of equality.” Or, Thomas Nagel writes that “the requirement of impartiality can take various forms, but it usually involves treating or counting everyone equally in some respect – according them all the same rights, or counting their good or their welfare or some aspect of it the same in determining what would be a desirable result or a permissible course of action.” It is, of course, impossible to survey each and every potentially relevant

---

account. Equally impossible is making a definitive statement about the relationship between equality and impartiality in the liberal tradition as a whole; the tradition is simply too vast and varied. In focusing selectively on Smith, Mill, Rawls, and Habermas, I have tried to invoke some prominent thinkers in order to lend credence to my suggestion that impartiality in liberal theory – including both distributive and justificatory impartiality – is predominantly motivated by a normative commitment to equality.49

While it is illuminating to discern the role that equality plays in the diverse arguments of these authors, the differences between them may be more obfuscating than clarifying. What matters, after all, is not so much the specifics of any argument advanced by one or another thinker, but rather the conceptual trope that links equality and impartiality across the better part of an entire tradition of thought. We will do well, though, to identify that trope at some level of precision, in order that we may better understand it, identify its limitations, and appreciate it in relation to stability-based account of impartiality that we will explore in subsequent chapters. Toward this end, this section will aim to construct a stripped-down version of the egalitarian account of impartiality. Specifically, we will consider a formal argument showing how a normative commitment to equality motivates the two theses of impartiality. This will serve as something of an ideal-typical articulation of the egalitarian approach, capturing the critical essence of this understanding of impartiality.

Before proceeding, two caveats. First, I do not claim that the generic argument I will construct tracks perfectly on any one of the arguments surveyed above – much less all of them. As I showed, equality has been interpreted to require impartiality in various ways, so a generic version will inevitably miss or diverge from important features of these accounts. My aim is not

49 Additionally, see Alexa Zellentin, Liberal Neutrality: Treating Citizens as Free and Equal (Boston: De Gruyter, 2012).
to reconstruct those arguments, but to show precisely one way that equality and impartiality can be understood to relate.  

Second, it is not my intention to defend the egalitarian account here. While I shall strive to articulate a coherent line of argument, I am not committed to its validity, nor to the truth of its claims. In the final part of this chapter, I will highlight a weakness of the egalitarian account of impartiality, though this focuses not on the structure of the account itself, but on the coherence of its prescriptive conclusions. At present, my aim is merely to make clear one version of the grounding for impartiality that predominates in liberal theory – namely the egalitarian account.

The first step in such an argument is to identify exactly what the premise of equality consists in. Is it something like a Hobbesian conception of natural equality in virtue of some universal characteristic (such as vulnerability to violent death)? Or is it more like a Lockean notion of moral equality as bearers of natural rights? These are, of course, fundamentally divergent ways of conceptualizing equality, and they generate different sets of political implications. But the thrust of a basic assumption of equality can be captured in abstract form that leaves undetermined the precise nature of that equality:

(1) Persons have equal political standing in virtue of possessing some certain characteristic(s) belonging to set $S$.

This may appear, at first glance, too formalistic a premise to have any force. An awful lot depends, for instance, on what characteristics (belonging to $S$) establish equal political standing.

---

50 I have taken equality to be a premise from which impartiality is a derivative conclusion, but other relations are of course conceivable. See John Kane, “Justice, Impartiality, and Equality: Why the Concept of Justice Does Not Presume Equality,” *Political Theory* 24 (1996): 375–93, e.g., for the view that neither equality nor impartiality is more basic than the other. I address this view in more detail below.
But recall that the aim here is a generic argument, the specifics of which might be filled in in various ways. Hobbesian and Lockean formulations of this premise, for instance, would identify distinct sets of characteristics as belonging to the determinative set $S$. The thrust of the premise is simply that equality is established by virtue of one being a certain kind of being, whether that being is ‘a human,’ ‘a rational animal,’ ‘a human vulnerable to violent death,’ or any other formulation. Any being which lacks the requisite characteristic cannot be considered an equal in terms of political standing.

Note that one thing this implies is that an egalitarian account need not be cosmopolitan. That is, equality may be taken to inhere in persons by virtue of sets of characteristics that are more or less exclusive. For instance, premise (1) is equally valid whether $S$ includes only the characteristic of ‘human’ or a more exclusive set of characteristics, such as {‘human’ and ‘rational actor’} or {‘human,’ ‘male,’ ‘white,’ and ‘land-owner.’} Views of the latter sort may be substantively objectionable, but they may be internally consistent insofar as they simply raise the bar for membership in the class of beings considered as equals. Liberal theories tend to set that bar rather low, but there is nothing in the structure of the basic premise of equality that requires universal participation in equal political standing.

From premise (1), it follows that characteristics other than those belonging to $S$ do not affect one’s standing as a political agent. Thus:

(2) Characteristics not belonging to $S$ do not bear on political standing, and are therefore politically irrelevant.
“Politically irrelevant” must be understood correctly. It is possible to affirm that there are no personal characteristics that are irrelevant to politics. Tall persons, for instance, are paid more, on average, than short persons; economic and class status are of obvious political relevance. A person’s temperament, physical attractiveness, or professional sports allegiances (for example) may be politically “relevant” in the sense that they may influence their or others’ political actions. So perhaps every characteristic of every individual is politically relevant in some way.

The claim of (2), though, is much narrower. It merely specifies that because an egalitarian political view takes political standing to be established in virtue of some characteristics, political standing is therefore indifferent to all other characteristics. The vast majority of personal traits will fall into the category of political irrelevance: left- vs. right-handedness, preference for coffee vs. tea, natural singing ability, and so on. This, for our purposes, is the meaning of political irrelevance.

Premises (1) and (2), in combination, specify a normative definition of equality. They define in what equality consists and what is irrelevant to one’s standing as (political) equal. But to move the argument forward, it is necessary to introduce some premise concerning the normative implications of equality. The basic thrust of that normative premise must be along the lines of respecting or according with the equality established by (1) and (2). This idea can be expressed in generic form as:

(3) By virtue of their equal political standing, persons have equal claims to political consideration (of their interests/preferences/rights, etc.).
This supplies content to the notion of “political standing.” To be an agent of political standing equal to another is to enjoy an equal claim on the way that one is regarded by political arrangements. Different theories of politics will spell out differently what it means for a political agent to be regarded appropriately: to serve her interests, to satisfy her preferences, to respect her rights, and so on. What these have in common is the commitment that persons merit equal consideration in the determination of political arrangements. This premise links the fact of equality to its political consequences.

The next step toward a prescriptive conclusion is to posit some function of political arrangements – that is, laws and institutions:

(4) Political arrangements serve to distribute benefits and burdens among members of a political society.

“Benefits and burdens” is a deliberately vague phrase that might encompass utility increases/diminutions, rights protections/violations, or preference satisfaction/dissatisfaction. Whatever metric of benefit and burden is taken to be central, at least one job of politics is to specify arrangements that distribute those benefits and burdens among the members of a political society. Politics may serve various ends, but one of these is surely its distributive function.51

Premises (1)-(4), taken together, support the following conclusion:

(5) Political arrangements ought to distribute benefits and burdens impartially with respect to politically irrelevant differences among persons (Distributive Impartiality).

---

51 For the sake of simplicity, I assume that any other political functions exist harmoniously with the distributive one, setting aside any cases where a distributive principle is superseded or overridden by some other political principle.
People who are equal by virtue of possessing the requisite characteristics have equal claims on the political distribution of benefits and burdens, such that politically irrelevant differences ought not to influence that distribution. In other words, the political distribution of benefits and burdens must be impartial with respect to politically irrelevant differences. For Rawls, those differences include such facts as social position, race, gender – all those features obscured by the veil of ignorance. Principles of justice ought to be impartial with respect to the distribution of those qualities among citizens. For utilitarians such as Mill and Sidgwick, the distribution of utility ought to be similarly impartial, insofar as there is no prima facie reason to favor one person’s satisfaction over that of another. Smith’s impartial spectator serves to remove from consideration morally (and politically) irrelevant facts of personal attachment and interest. For each of these, impartiality in the distribution of benefits and burdens is a requirement of recognizing and treating each person in accordance with their status as equals.

We have touched several times on the observation that recognition and treatment of each person as an equal need not entail uniform or materially identical treatment. Rawls’s difference principle demonstrates this well: from a position of equality in the original position, parties select principles of justice that permit inequalities in the resulting society, with the condition that those inequalities facilitate absolute increases in the well-being of the least well-off members of the society. That citizens may end up in relationships of inequality – even in terms of important goods like money – poses no necessary conflict with their fundamental equality qua citizens. Dworkin has captured this distinction neatly as the difference between equal treatment and treatment as an equal. The former entails a “right to an equal distribution of some opportunity or resource or burden,” while the latter is more basic: “the right … to be treated with the same
respect and concern as anyone else.”\textsuperscript{52} Sometimes, treatment as an equal will require equal
treatment, but not always, and equal consideration – premise (3) – need not be assumed to entail
strictly egalitarian political prescriptions.

Failure to distinguish between equal treatment and treatment as an equal underwrites at
least one critical take on the assumed relationship between impartiality and equality. In a 1996
article in \textit{Political Theory} titled “Justice, Impartiality, and Equality,” John Kane disputes the
centrality of equality as a motivating premise for impartial justice. In particular, Kane takes aim
at Amartya Sen’s claim that the very idea of justice rests on a principle of “presumptive
equality,” where equality is the presumptive norm, and any deviations from equal treatment fall
under a burden of justification.\textsuperscript{53} Such justification must rest on an appeal to some consideration
that justifies unequal treatment: either some independent priority that overrides concern for
equality or some quality in respect of which persons can be judged to merit differential
treatment. Kane agrees with Sen that the connection between impartial justice and equality has
become “deeply embedded in the modern liberal conscience,” but, \textit{contra} Sen, he thinks this
connection is mistaken. On Kane’s view, justice is concerned with \textit{proportionality} in the
Aristotelian sense; he defines the principle as follows: “the more one has of a certain property X,
the more one deserves of whatever treatment Y is relevant to it.”\textsuperscript{54} This is a purely formal
principle that says nothing about the basic equality or inequality of persons. It is fully compatible
with the above premises regarding politically relevant and irrelevant differences among persons,
but does not itself rely on any assumption of equality. Thus, Kane concludes, there is no
necessary connection between an idea of justice and an assumption of equality.

\textsuperscript{54} Kane, “Justice, Impartiality, and Equality,” 378.
Sen issued a response to Kane in that same issue of *Political Theory*, and a key element of his rebuttal turns on the distinction between equal treatment and treatment as an equal. The proportionality principle, while it is not an overtly egalitarian principle, does at a basic level reflect a commitment to equality. It prescribes differential treatment of persons in virtue of the characteristics they possess, but it recognizes and responds to those characteristics without preference or bias in regard to the persons who possess them. As Sen puts it, “the Y that is ‘due’ (to use Kane’s terminology) to a person is supervenient on his or her X, and not on who the person is (or on any other characteristic of this person).”

55 No person possessing quality X has any greater or lesser (*pro tanto*) claim to the corresponding good Y than any other person similarly situated. Kane is correct that equal treatment requires substantive justification – it is not a necessary requirement of justice – but the impartial principle of justice he invokes does itself reflect a basic commitment to equality in virtue of its treatment of persons *as equals*. Thomas Nagel captures the basic role of equality in motivating impartiality nicely:

“The impartial attitude is, I believe, strongly egalitarian both in itself and in its implications. As I have said, it comes from our capacity to take up a point of view which abstracts from who we are, but which appreciates fully and takes to heart the value of every person’s life and welfare. We put ourselves in each person’s shoes and take as our preliminary guide to the value we assign to what happens to him the value which it has from his point of view. This gives to each person’s well-being very great importance, and from the impersonal standpoint everyone’s primary importance, leaving aside his effect on the welfare of others, is the same.”

So much for the egalitarian basis for distributive impartiality; on now to justificatory impartiality. I suggested the outlines of this argument above, so it remains only to make it explicit. The first step is to introduce a companion premise to equality: autonomy.

(6) Persons enjoy a pre-political claim against infringements on their autonomy.

To be sure, not all will affirm this claim, nor will all those who affirm it agree on the meaning of autonomy or on what constitutes an infringement thereof. But something like (6) can be endorsed on a number of grounds. On its own, it entails no particular metaphysical commitments, and it can support a range of prescriptive conclusions. But it does require that acts of coercion (which infringe on the autonomy of individuals) meet some standard of justification that can override the _pro tanto_ claim against such coercion. One way this standard might be met is by an appeal to authority: such-and-such coercion is justified by virtue of being enacted or mandated by some person with the legitimate authority to do so, where legitimacy is established by a rule of recognition that confers some kind of authority on some individual(s). The town curfew is legitimate by virtue of having been ordered by the police chief. The authority of that officer, however, depends on a web of prior coercive arrangements: she does not possess the authority _qua_ herself, but rather as an occupant of a well-defined political role. As an individual person, she has no legitimate authority to impose coercion on another. Generally:

(7) By virtue of (1), no person has, by default, legitimate authority to coerce another person of equal political standing by political means.
Authority may be granted or achieved by some means (democratic procedure, direct delegation, etc.), but no person possesses coercive authority over others by default, because persons are of equal political standing by virtue of their sharing in the requisite characteristic(s). In the absence of some natural authority, coercion may be legitimate alternatively by virtue of justification: supplying reasons in support of the coercive arrangement that the person subject to the coercion could herself endorse. Kant and Rousseau are the clear standard-bearers for this way of thinking about the compatibility of freedom and coercion. (I won’t engage them here, though I do take up Rousseau and his idea of the general will in Chapter 4.) This leads us to the next premise:

(8) Political arrangements must be justified on the basis of reasons that the person(s) subject to their requirements could accept.

From here, we proceed via the fact of pluralism to the conclusion of Justificatory Impartiality.

(9) For any comprehensive doctrine C affirmed by some citizen(s), there will be some other citizen(s) whose own comprehensive doctrine(s) reject the truth of C.

(10) A regime that depends for its justification on some particular comprehensive doctrine C will fail to win endorsement from those citizens whose comprehensive doctrines reject the truth of C.

57 I am setting aside the issue of paternalistic care for children, the elderly, the severely disabled, etc.
This means that, in order for a regime to be justified to all the citizens who live under it:

(11) Political arrangements ought to be supported by a justification that is impartial with regard to comprehensive doctrines. (*Justificatory Impartiality*).

Premise (11) follows from (10) in the following manner: if a justificatory account relying on one comprehensive doctrine fails to be justified because it conflicts with some citizen’s alternative comprehensive doctrine, then the satisfaction of premise (7) requires that political arrangements be supported by justifications that depend on the truth of *no* particular comprehensive doctrine.

The above argument demonstrates how the general principles of distributive and justificatory impartiality can be supported by an argument proceeding from the initial premise of equality. Proponents of liberal impartiality tend to hold that treating persons appropriately in light of their status as equals requires that benefits and burdens be distributed without reference to politically irrelevant differences among them, and that the laws and institutions that govern them and their relations be justified on grounds that are impartial with respect to the various comprehensive doctrines that citizens affirm.

In the next section I wish to highlight one limitation of the egalitarian basis of impartiality in regard to the case of religious exemptions from generally-applicable laws. Specifically, equality can ground the distributive and justificatory components of impartiality, but the argument from equality to impartiality is overly formal. In particular, it does not say enough about what differences among persons are politically relevant, and thus merit
accommodation. As I will show, equality can be marshaled in support of opposing conclusions on religious exemptions, depending how the content of the egalitarian account is filled in.

1.5 The Indeterminacy of Equality: Religious Exemptions

So far in this chapter, I have highlighted the prominent conceptual closeness of equality and impartiality in liberal theory, and I have constructed one version of an argument, in highly generic form, that grounds the principles of distributive and justificatory impartiality in a normative commitment to equality. My overarching aim in this project is to show that the egalitarian basis of impartiality is not the only possible one. The end of stability amid conditions of pluralism, I will argue, also provides support for the normative propositions of distributive and justificatory impartiality. I intend this argument as a positive contribution to liberal theory, highlighting neglected lines of argument that are available to shed new light on some of the difficult issues facing liberal societies and liberal theory today. In this spirit, it is not my project to advance any thoroughgoing critique of egalitarian arguments for impartiality. I have neither need, desire, nor grounds to categorically reject this basis of impartiality.

That being said, my aim here is to highlight one limitation of the egalitarian account of impartiality, not so much in order to discredit the egalitarian account, but rather to lend credence to my suggestion that an alternative justification for impartiality represents a helpful contribution. The limitation I have in mind concerns the ability of equality to yield determinate conclusions on the permissibility and scope of religious exemptions from generally applicable laws. I introduced the challenge of religious exemptions in the Introduction, and suggested there that they represent an area of dispute in both jurisprudential and philosophical discourse. My aim here is to show that a commitment to equality is insufficient to resolve this matter, as equality
can be marshalled in support of normative prescriptions both against and in favor of religious exemptions. Equality is, ultimately, indeterminate on whether and which religious exemptions ought to be granted.

In the generic argument offered above, a foundational premise of equality grounded the conclusion that political arrangements should distribute benefits and burdens impartially with respect to differences among citizens that do not bear on their respective standings as political agents. Among the characteristics that would ordinarily be regarded as politically irrelevant, in this sense, is a citizen’s religious beliefs. One’s religious identity, belief, or practice has no bearing on one’s standing as a political agent – or at least most liberals would affirm as much. Thus, it would follow that equality requires that laws and political institutions distribute benefits and burdens without reference to religious differences among citizens. To grant exemptions based on citizens’ particular religious commitments would be to show partiality toward one or another religious belief – or perhaps to religious belief itself, over against non-belief.58 According to the argument from equality, religious differences among citizens are the kind of consideration that political arrangements ought to “shun.”

This is not a consensus conclusion, however. Rather, as we shall see momentarily, the norm of equality has been invoked on both sides of this issue – both against and in support of accommodating religious difference. A brief examination of three sample views will demonstrate this ambivalence. As we proceed, note that the arguments we will be considering do not track the generic argument exactly, and may diverge in some important respects. That said, my goal is to show that they share in many of the basic premises of the generic argument, while nevertheless arriving at divergent prescriptive conclusions.

Brian Barry

Brian Barry, in his 2001 book *Culture & Equality*, offers a critique of religious and cultural accommodation that tracks the major components of the generic argument from equality. In short, he argues that accommodation of religious and cultural difference runs afoul of a commitment to equality, and that exemptions should be ruled out, or at least minimized, for this reason.

Barry argues that the idea of impartial treatment under the law is a fundamental element of the liberal conception of citizenship. That conception takes equality as bedrock, and holds that the law should obligate citizens uniformly. This means that it should be indifferent to traits such as religious identity or group membership. The liberal tradition has held, in Barry’s words, that “there should be only one status of citizen (no estates or castes), so that everybody enjoys the same legal and political rights. These rights should be assigned to individual citizens, with no special rights (or disabilities) accorded to some and not others on the basis of group membership.”

Barry notes that this commitment did not arise in a context of homogeneity—which would deflate the significance of the law’s indifference to group identity— but rather was devised as a solution to pluralist conflict among religious factions in 16th and 17th century Europe. (This claim relates closely to arguments that we will consider closely in subsequent chapters.) Barry’s point is that uniform legal treatment of persons across difference is no minor or accidental feature of liberalism, but is rather one of its fundamental innovations. “[L]iberal principles,” he writes, “can make a moral appeal as a fair way of solving conflict, because they offer the parties equal treatment.”

---

60 Ibid., 25.
Like many egalitarians, Barry does not assume that equal treatment requires strict equality in all respects. Rather, equal treatment means that persons must be afforded the same opportunities: “From an egalitarian liberal standpoint, what matters are equal opportunities.”\(^61\) Such a statement reflects an awareness that individuals’ opportunities within civil society are never unlimited: no matter the political arrangements, some opportunities will inevitably be foreclosed. This is, according to Barry, “the essence of law”: “the protection of some interests at the expense of others when they come into conflict.”\(^62\) For instance, he notes, the interests of children conflict with the interests of potential child abusers, and decent societies everywhere work to foreclose the opportunity of the abuser to act on this interest. That some opportunities will be unavailable or access to them significantly burdened in liberal society is not a problem, so long as access to those opportunities is burdened equally for all persons.

Of course, people differ in the opportunities that they desire or are committed to pursue, so the impact of opportunity limits will vary across persons. To continue the example from above, a person with no desire to abuse children will be effectively unconstrained by legal protections for child well-being, while persons who have such desires will be precluded from carrying out the actions they otherwise might. Similarly, in the U.S., people with a taste for wine are legally permitted to explore that interest, while people with an analogous taste for marijuana are, in general, not. Neither of these legal arrangements violates the principle of equal opportunity. Rather, as Barry describes, “the rules define a choice set, which is the same for everybody; within that choice set people pick a particular course of action by deciding what is best calculated to satisfy their underlying preferences for outcomes.”\(^63\) In other words, if an action falls within the choice set defined by the law, then each individual enjoys equal opportunity to

\(^{61}\) Ibid., 32.
\(^{62}\) Ibid., 34.
\(^{63}\) Ibid., 32.
engage in that action. The fact that people may differ in the level of satisfaction that they are able to derive from the available choice set does not mean that equality of opportunity has been violated.

Barry points to a well-known debate in the British parliament over whether Sikhs who are religiously obligated to wear turbans ought to be exempted from laws requiring motorcyclists to wear helmets.64 Some have suggested that such a law, without an exemption, is tantamount to a law stating that “Sikhs may not ride motorcycles.”65 Barry, however, rejects the notion that a helmet law that applies universally is equivalent to a facially discriminatory law. “Once again,” he writes, “we must insist on the crucial difference between a denial of equal opportunities to some group (for example, a law forbidding Sikhs to ride motorcycles) and a choice that some people make out of that from a set of equal opportunities (for example, a choice not to ride a motorcycle) as a result of certain beliefs.”66 The Sikh enjoys exactly the same opportunities as everyone else; he simply prefers not to take on the cost of riding the motorcycle (removing his turban, namely), though he is perfectly free to do so. We all routinely make calculations of this sort, declining to pursue certain opportunities for reasons of preference, cost, or belief. That some will elect to forego some of the opportunities that are legally afforded them does not imply any violation of equality of opportunity.

In the terms of the generic argument above, Barry holds that the equality standing of persons as political agents is paramount in liberal theory. Laws define legal choice sets, and they ought to do this in a manner that is uniform, not variable with respect to persons’ cultural identities or religious beliefs. To permit religious exemptions is to create inequalities of

64 Ibid., 44–49.
66 Barry, Culture and Equality, 45.
citizenship status. Granting some people – but not others – exemptions from generally-applicable laws creates two classes of persons: those who are obligated by the law and those who are not. Those who receive no exemption remain constrained by the law, with no right to either engage in the conduct forbidden by the law or to refrain from the conduct mandated by it. Those with the exemption, in contrast, enjoy a right that others do not.67 This unequal distribution of the benefits and burdens of political order is in violation of the basic liberal commitment to equality. It offends, Barry argues, against the liberal conception of citizenship, by creating distinct and unequal classes of citizens within a society.

Barry does not go so far as to say that exemptions are never permissible. He admits that in some cases there may be a sufficient pragmatic case for granting exemptions. Sometimes, selective exemptions will, all things considered, be a more satisfactory solution than either insisting on universal enforcement or, alternatively, repealing the law whole-hog. In evaluating cases where exemptions are requested, we do well to consider all the potential effects of enforcement, repeal, or enforcement. Sometimes, Barry concludes, “it is preferable to give up on consistency than abandon the advantages of the present legislation.”68 However, exemptions that

67 Simon Caney argues that exemptions do not necessarily create unequal rights: in some cases an exceptional benefit is balanced by some alternative cost. For example, he notes, pacifists can be exempted from military service, but must take on medical work instead. See Simon Caney, “Equal Treatment, Exceptions and Cultural Diversity,” in *Multiculturalism Reconsidered: Culture and Equality and Its Critics*, ed. Paul Kelly (Malden, MA: Polity Press, 2002), 88. This seems to miss the facts that a) the alternative cost may be appreciably less burdensome than the original exemption (e.g., medical work is less risky than military service), and b) simply having the option to choose between obeying the law and taking advantage of an exemption is a liberty that non-exempted citizens do not enjoy. It also might appear that legally codified exemptions avoid the problem of unequal status, since everyone is subject to the same legal rules. For example, a law could require that “All whose religious beliefs do not require the constant wearing of headgear that would make the simultaneous wearing of a helmet impossible are required to wear helmets when riding a motorcycle.” This approach avoids granting some a right to break the law by building conditionality into the law itself. This is only a technical dodge, though. The function of such a law is obviously to afford special treatment to some class of persons. Claiming that it creates no inequalities of rights is tantamount to suggesting that the tax code taxes all income uniformly because the tax code applies to all income, regardless of the tax rate it imposes. Such a formulation obscures the fact that an exemption clearly establishes a dispensation from a legal obligation, even if the exemption is written into the law itself. (Thanks to Henry Richardson for pressing this point, and for the formulation of the sample statute.)

68 Barry, *Culture and Equality*, 51.
are justified in this case do not prove any rule; rather, “they are anomalies to be tolerated because the cure would be worse than the disease.” In general, Barry holds, exemptions are not the preferred strategy. Rather, when an exemption appears desirable, it is likely that the law itself needs to be repealed, or at least re-formulated. For instance, that allowing Rastafarians to smoke marijuana for religious purposes strikes many as a reasonable accommodation speaks not in favor of an exemption, but rather in favor of repealing the prohibition on consumption of cannabis for everyone.69 In other cases, the reasons supporting the law will be sufficiently important to rule out exemptions. Whatever the outcome, this is Barry’s preferred approach to reconciling the commitments of believers with the obligations of law. Rather than carve out exemptions for designated special groups, we ought to “work out some less restrictive alternative form of the law that would adequately meet the objectives of the original one while offering the members of the religious or cultural minority whatever is most important to them. This avoids the invidiousness of having different rules for different people in the same society.”70

Bhikhu Parekh

Barry’s objectivist conception of opportunity, where opportunity is defined by the formal choice set defined for each person, regardless of each’s ability to practically or effectively exercise those choices, begs for an objection. Bhikhu Parekh has advanced a defense of cultural

69 Ibid., 40.
70 Ibid., 39. For a critical exploration of Barry’s arguments in Culture & Equality, see Paul Kelly, ed., Multiculturalism Reconsidered: Culture and Equality and Its Critics (Malden, MA: Polity Press, 2002). Additionally, Maria Paola Ferretti, “Exemptions for Whom? On the Relevant Focus of Egalitarian Concern,” Res Publica 15 (2009): 269–87, has argued that authors such as Kymlicka, Parekh, and Barry have focused on identifying the criteria that qualify practices for exemptions, to the neglect of the (in her view) more important question of who should be entitled to exemptions.
and religious accommodation on grounds very similar to Barry’s, though his reliance of a different conception of opportunity leads him to starkly different conclusions.\textsuperscript{71}

Parekh’s basic claim is that laws that disregard differences among persons fail to treat those persons equally. He writes that “treating human beings equally requires us to take into account both their similarities and differences.”\textsuperscript{72} People are not identical, and some of the differences among them bear significantly on important political questions. There are, Parekh says, “legitimate and relevant” differences between persons, and an egalitarian political theory must account for these. In practical terms, this means that some accommodations of cultural practice and religious belief are appropriate. Parekh cites another example involving difficulties in assimilation faced by adherents of the Sikh faith: a Sikh father who wishes to send his child, who wears a turban, to a school where the wearing of turbans is banned. There is no overt constraint on the father sending child to such a school, but it would involve forcing the child either to remove the turban or to face constant disciplinary action. In light of these burdens, “for all practical purposes [the father’s option to send his child to this school] is closed to him.”\textsuperscript{73} For the Sikh family to have an effective opportunity to attend the school in question, some accommodation of their commitment to wearing that particular garment must be made.

The discussion so far may give the impression that Parekh disputes premises (1) and (2) from the generic account, which suppose that characteristics other than those which establish persons’ equal standing as political agents were politically irrelevant, and to be shunned in political matters. Parekh, in contrast, explicitly names some differences as politically relevant.

\textsuperscript{72} Bhikhu C. Parekh, \textit{Rethinking Multiculturalism: Cultural Diversity and Political Theory} (Cambridge, MA: Harvard University Press, 2000), 240.
\textsuperscript{73} Ibid., 241.
The Sikh student’s religious identity is relevant to the determination of the dress-code regulations that will apply to him. It might seem that Parekh’s account does not so much suppose equality as a foundational premise, but rather aims for the guarantee of equality in terms of practical outcomes.

A closer look, however, shows Parekh’s departure from Barry’s approach to religious accommodation is actually fully consistent with the generic account, and differs only in its interpretation of one of the later premises supporting Distributive Impartiality. The premise of basic equality in terms of political standing is implicitly endorsed by Parekh’s account. That he considers the practical inaccessibility of opportunities for the Sikh family to be a problem reflects his assumption of the Sikh’s equal claim to political consideration. Religious belief does not determine one’s basic political standing, on Parekh’s view, or else there would be no obvious problem with the Sikh having access to a different set of opportunities than non-Sikhs.

The problem – and here lies the crux of Parekh’s disagreement with Barry – is that a political distribution of benefits and burdens that is not attentive to some differences among persons will result in systemic and predictable disadvantage to some persons.74 For the Sikh family to have meaningful access to educational opportunity – in other words, for them to effectively receive that benefit – some accommodation of their commitment to wearing that particular garment must be made. Parekh differs from Barry in conceptualizing opportunity as more than a formal concept. In contrast to Barry’s objectivist conception, Parekh takes opportunity to be “a subject-dependent concept in the sense that a facility, a resource or a course of action is only a mute and passive possibility and not an opportunity for an individual if she lacks the capacity, the cultural disposition, or the necessary cultural knowledge to take advantage

---

74 I specify that this disadvantage is “systemic and predictable” to distinguish from those variances in access to opportunity that will inevitably result from contingent and non-foreseeable factors. That Sikhs will be burdened by rules prohibiting turban-wearing is not a matter of chance, nor is it impossible to see in advance.
of it.”

In order for a person to truly have access to an opportunity, she must be actually able to take advantage of it. The upshot of this understanding of opportunity is that some political distributions of benefits and burdens that are ostensibly indifferent to religious and cultural identity can result in effectively partial distributions, in that the distribution will vary depending on religious and cultural identity. A legal order that does not accommodate for religious belief is therefore eminently partial, insofar as it confers different benefits on adherents of different religions.

Where Parekh and Barry part ways, then, is over their interpretation of premise (4): “Political arrangements serve to distribute benefits and burdens among members of a political society.” Specifically, the two authors rely on different understandings of exactly what benefits are supposed to be distributed by political arrangements. For Barry, the relevant distribuendum is formal access to opportunity; for Parekh, it is substantive access to opportunity. On Parekh’s view, in order to distribute access to opportunity in a way that does not depend upon each citizen’s religious identity, some accommodation is necessary. He writes, “Equal rights do not mean identical rights, for individuals with different cultural backgrounds and needs might require different rights to enjoy equally in respect of whatever happens to be the content of their rights.” A distribution of rights that accords with persons’ basic standing as political equals, and that does render different rights to persons of different beliefs, may require some facially “unequal” treatment in the form of accommodation.

---

75 Parekh, Rethinking Multiculturalism, 241.
76 Parekh’s view has affinities with the capabilities approach advocated by Sen and Nussbaum, which takes into account the fact that persons will have different abilities to convert resources into welfare, or capabilities. See Martha C. Nussbaum and Amartya Sen, eds., The Quality of Life (Oxford: Clarendon Press, 1993); Sen, Inequality Reexamined, 85.
77 Parekh, Rethinking Multiculturalism, 240.
One of the most prominent equality-based accounts in support of accommodation of religious and cultural difference (including religious exemptions) has been advanced by Will Kymlicka. Kymlicka’s argument falls among a family of egalitarian views that have come to be known by the label of ‘luck egalitarianism.’ The driving intuition of luck egalitarianism is that simple fortune — the fact that A, through no fault or merit of her own, is better off than B — is not a sufficient justification for inequality. G.A. Cohen, a primary exponent of the luck egalitarian view, writes that “a large part of the fundamental egalitarian aim is to extinguish the influence of brute luck on distribution.” By focusing on brute luck, Cohen means to distinguish from what he calls “option luck,” which stems from gambles that a person either freely chose or could have avoided. The desire to eliminate advantages and disadvantages conferred by brute luck has characterized much of contemporary theorizing about equality. For instance, Rawls argues in an early section of *Theory of Justice* that “since inequalities of birth and natural endowment are undeserved, these inequalities are somehow to be compensated for.”

Kymlicka’s approach has been described as “luck multiculturalism,” in the sense that it is concerned with the unequal advantages or disadvantages that may result from the chance feature of a person’s life that they may be born into a minority cultural group or belief tradition.

---

These people may face what Kymlicka describes as “a disadvantage with respect to cultural membership.”

The specific disadvantage arises from difficulties that members of minority cultural group are likely to face in maintaining their cultural identities and traditions in the face of laws shaped by majority cultural groups. Kymlicka writes that members of minority cultures face the possibility that “the viability of their societal cultures may be undermined by economic and political decisions made by the majority.” The reason for this is that “the majority nation will always have its language and societal culture supported, and will have the legislative power to protect its interests in culture-affecting decisions.” For members of minority groups to protect their interests and maintain their cultures requires of extra resources – which members of majority cultures do not have to spend in order to achieve the same good. Thus, to belong to a minority culture is to bear burdens that members of the majority do not. Kymlicka writes that “this is a significant inequality which, if not addressed, becomes a serious injustice.”

Avoiding this injustice, then, requires that members of minority cultures be compensated in some manner for the unequal burdens they bear. After all, membership in cultural groups is very rarely a matter of choice. For all but the very few, membership is determined by birth. It is

---

83 Kymlicka, Multicultural Citizenship, 109–110.
84 Ibid., 113.
85 Ibid., 109.
86 Actually, some such injustices should be remedied by repealing or amending the majority’s inconsiderate or oppressive political decisions. At times, Kymlicka seems to assume minimal agency on the part of the majority, taking it for granted that they will thoughtlessly take actions that impose avoidable burdens on the minority. For some such decisions, exemptions are an insufficiently radical fix; rather than carving out exceptions to an unjust rule, the rule itself ought to be fixed or eliminated altogether. That said, some of the burdens that Kymlicka describes will arise from majority decisions that are not themselves unjust. For instance, the federal calendar can accommodate only so many official holidays, and it is reasonable to prioritize those that are observed by a sizeable portion of the population. That this imposes a certain burden on members of minority religions whose holidays are not included does not imply that the calendar is unjust. Some burdens are simply unavoidable consequences of minority status.
87 Barry takes a different view of religious identity, likening religious beliefs to “expensive tastes,” which people can reasonably be held responsible for cultivating, if not choosing outright. Barry, Culture and Equality, 35–37.
therefore a matter of brute luck that some persons face unequal burdens in maintaining important aspects of their cultural identity. The answer, Kymlicka argues, is to grant special group rights to help support minority cultural structures that stand under threat from the decisions of the majority group. For example, he suggests that Muslims and Jews, whose days of religious observance are often not recognized by public work calendars, should be granted an exemption from requirements to work on those days. Or, Sikhs and Orthodox Jews should be exempted from police or military dress codes that forbid their culturally-specified headgear. In both of these cases, the general standards have been constructed with Christian practices in mind (workplaces often close on Sundays, wedding rings are not prohibited by police dress codes). Affording exemptions to members of minority cultures represents an effort to “compensate for unequal circumstance which put the members of minority cultures at a systemic disadvantage in the cultural market-place, regardless of their personal choices in life.” Just as Parekh defended accommodation as a matter of treating persons as equals, Kymlicka contends that “This is one of many areas in which true equality requires not identical treatment, but rather differential treatment in order to accommodate differential needs.”

This argument begs the question of why maintenance of a “societal culture” is to be valued more highly than many other kinds of projects that citizens might pursue. For instance, members of wine-tasting clubs will find it harder to maintain their group identity and traditions in jurisdictions where alcohol sales are prohibited. But from this fact it doesn’t follow that wine-tasters deserve exemptions from blue laws. So why afford such deference for maintenance of minority cultures? The answer is that Kymlicka assumes that membership in a group is a core human good, one that is necessary or the realization of individual autonomy. Cultural membership provides what Kymlicka calls a “context for choice.” He writes, “We decide how to

---

88 Kymlicka, Multicultural Citizenship, 113.
live our lives by situating ourselves within these cultural narratives, by adopting roles that have struck us as worthwhile ones, as one worth living.”89 I will not rehearse Kymlicka’s argument in full detail here, but the bottom line is that autonomous choice is a necessary condition for a good life, and belonging to a cultural group provides the context and conditions for autonomous choice. In short, access to cultural membership is a sufficiently important good that political arrangements should be determined with the distribution of this good (or of burdens to persons’ realization of this good) in mind.

As for Parekh, the determining component of Kylicka’s argument lies in the content he supplies to the fourth premise. His account implicitly affirms premises (1), (2), and (3): he assumes the basic equal political standing of individuals and their equal claim to political consideration regardless of characteristics such as cultural identity and religious belief. In some cases, making good on these commitments will require attention and sensitivity to certain differences among citizens, in order to avoid a distribution of the good of cultural membership that depends on one’s religious or cultural identity.

Examining these three egalitarian approaches to religious accommodation – Barry’s, Parekh’s, and Kymlicka’s – reveals the profound indeterminacy of religion on this question.90 Each of the three affirms the basic premise of equality, taking individuals to enjoy equal political standing regardless of “irrelevant” considerations, including religious belief. For none of them does religion alter a person’s political standing vis-à-vis his fellow citizens. They each affirm the

89 Kymlicka, Liberalism, Community, and Culture, 165.
90 In addition to the arguments we have considered here, there are other ways to affirm exemptions on egalitarian grounds. For instance, Paul Bou-Habib, “A Theory of Religious Accommodation,” Journal of Applied Philosophy 23 (2006): 109–26, argues in support of religious exemptions on the basis that freedom of religious exercise is required for persons to enjoy the basic good of integrity, which is in turn a necessary condition of equal opportunity for well-being. See also Quong, “Cultural Exemptions, Expensive Tastes, and Equal Opportunities”; Martha C. Nussbaum, Liberty of Conscience: In Defense of America’s Tradition of Religious Equality (New York: Basic Books, 2008); Tariq Modood, Multiculturalism: A Civic Idea (Malden, MA: Polity Press, 2007).
notion that persons have equal claims to political consideration. But despite this shared starting point, Barry’s conclusions regarding the desirability of religions exemptions are starkly opposed to those of Parekh and Kymlicka. What accounts for this divergence? The fact is that the prescriptive implications of the egalitarian approach to religious accommodation depend significantly on details of the argument that have nothing to do with equality as such. The factor that drives Barry to disfavor exemptions and Parekh and Kymlicka to favor them lies in the way each author gives content to premise (4), regarding the benefits and burdens that political arrangements are to distribute. Barry takes that distribution to consist in the definition of a legal choice set that applies equally for all citizens; Parekh understands it to consist in the provision of substantive access to opportunities; Kymlicka includes in that distribution the ability of citizens to maintain membership in a context of autonomous choice. The bottom line is that a commitment to equality itself cannot resolve the issue of religious exemptions. Equality can be invoked to yield wildly divergent conclusions depending on how one fills out other subsequent premises of an egalitarian argument. Equality itself is basically indeterminate on this matter.

**Legal Approaches**

I have so far focused on philosophical approaches to the question of religious exemptions, and have not waded deeply into jurisprudence on the topic. The primary reason for this is that in this chapter I am concerned to highlight the motivating role played by a normative commitment to equality, and it is difficult to isolate that variable in legal analyses of religious exemptions, which are influenced necessarily by myriad constitutional, statutory, and precedential considerations. In the U.S., for instance, rulings on religious exemptions begin not from an abstract commitment to equality, but rather from the guarantees of the First Amendment,
the interpretations of those guarantees established by prior Courts, and statutes (such as the Religious Freedom Restoration Act of 1993 or the Religious Land Use and Institutionalized Person Act) that have modified or augmented the rights guaranteed in the Constitution. To identify clearly what role equality plays in these considerations is not straightforward.

That is not to say that equality is not a factor in constitutional or statutory frameworks regulating religious exercise. Michael McConnell has shown that in the process of drafting the Free Exercise clause, the founders employed the language of equality frequently, and many early versions of the text included not merely a guarantee of right of free exercise, but protection of the “full and equal rights of conscience.” In the end, the First Amendment left out the language of “equal rights,” but the history of this text suggests that the rights enshrined in the constitution are importantly understood to be guaranteed equally. An equal guarantee of religious liberty requires that all persons be able to engage in their religious practice. As Kymlicka argued, adherents of majority religions will have less trouble doing so because laws will have been shaped with their commitments in mind. Adherents of minority religions, it might be thought, may require special accommodation to make good on their rights.

My point here is that the starting point of equality does not yield less ambivalent conclusions about religious exemptions inside the realm of legal analysis than it does in abstract theoretical analysis. This can be seen in a deceptively simple framework for religious accommodation put forward by Lawrence Sager and Christopher Eisgruber. They call their view “equal liberty,” and suggest that a bifocal commitment to equality and liberty can satisfactorily resolve disputes over religious exemptions. In short, their position is that a commitment to equal liberty requires that religion be regarded equally with analogous secular interests. When a secular

---

interest receives accommodation, then equality requires accommodating religious interests similarly. Alternatively, when a secular interest is trumped by the obligations of the law, then equality holds that the religious interest should be similarly overridden.

So, when two Muslim police offices in Newark objected to the department’s prohibition on facial hair, citing their religious obligation to grow beards, the fact that the police department already offered an exemption for officers suffering from folliculitis (a skin condition that makes shaving painful) led a U.S. Court of Appeals to rule that religious beliefs should also be accommodated. The same standard should also hold across different religious beliefs, so in Employment Division v. Smith, Sager and Eisgruber contend that an exemption to the peyote ban was justified on grounds that Oregon already offered exemptions from dry county regulations for church officials purchasing wine for sacramental purposes. To deny religious accommodation in these cases, would be to fall short of the standard of equal regard for secular and religious interests. If a non-religious concern receives an exemption, then a religious one should as well. Conversely, when non-religious actors are uniformly held to the standard of the law, then religious actors should be as well. For instance, churches should not receive dispensations from zoning regulations that secular organizations are required to follow. Sager and Eisgruber affirm Barry’s view of law as a domain in which interests conflict, noting that “we are regularly called upon to act in ways that we dislike. … We accept the imposition of these rules because our society – indeed any society – could not function without reciprocal sacrifices of this sort.” The important thing is that these sacrifices be distributed equally among religious and non-religious citizens alike.

93 Ibid., 92.
94 Ibid., 84.
The framework of “equal liberty” might seem to offer a straightforward application of the principle of liberty to the question of religious exemptions: accommodate religion neither more nor less than non-religion. The problem, as several commentators have pointed out, comes in determining which secular interests supply the relevant comparison when weighing a religious exemption.\textsuperscript{95} In the case of any particular law, there will usually be some considerations that warrant exemption from the law and some that do not. In the case of the Newark police officers, for example, officers could bypass the facial hair prohibition for medical reasons, but not for the purpose of following the model of an honored father, for example.\textsuperscript{96} Which of these non-religious interests is the appropriate benchmark case against which the religious interest in growing a beard should be compared? Nearly all regulations will be of sufficient import to rule out some reasons for exemption but not others, and the practical consequences of “equal liberty” depend dramatically on the construction of comparisons. Sager and Eisgruber provide no clear answer to this conundrum. The result is that equal liberty provides no reason to prefer equal accommodation – i.e., granting exemptions – over equal suppression – denying them. The theory is, as Thomas Berg has written, “entirely indeterminate – even incoherent – unless it somehow specifies which interests to compare to religion.”\textsuperscript{97}

\textbf{1.6 Conclusion}

What I have attempted to show in considering the egalitarian views of Barry, Parekh, and Kymlicka, along with the framework of “equal liberty” advocated by Sager and Eisgruber, is that


\textsuperscript{96} Berg, “Can Religious Liberty Be Protected as Equality?,” 1185.

\textsuperscript{97} Ibid., 1195.
while equality provides a fertile foundation for theorizing about pluralism, political obligation, and religious accommodation, it is too capacious a concept to generate determinate conclusions on how to resolve the question of religious exemptions. These various invocations of equality reveal its multivalent implications, some of which run directly counter to one another. On one hand, respecting the equality of subjects means uniformity in the laws that govern them. But, on the other, an equal distribution of important political goods may require special accommodation of minorities. On one hand, treating religion on equal footing with non-religion requires accommodating it neither more nor less than non-religion. But, on the other, aiming merely for equal regard cannot explain exactly what kind of treatment – accommodation or neglect – is appropriate for religion. Equality can plausibly support each of these conclusions, even though they cannot all be simultaneously true.

The generic account I constructed above showed how equality can be employed to ground the conclusions of both distributive and justificatory impartiality, but it also revealed how thoroughly dependent the specific content of those conclusions are on certain key assumptions, including the precise goods, benefits, and burdens that political arrangements are to distribute. Without relying on additional premises, the end of equality cannot supply determinate conclusions regarding whether political impartiality properly precludes or permits religious exemptions. I do not mean to suggest that it is impossible to argue from equality to some coherent position on exemptions – obviously, each of the authors surveyed in this section has done so. My contention is merely that doing so successfully requires introducing significant controversial content in addition to the basic idea of equality. One must, for example, take a position on the value of membership in a cultural community; or, one must take either a welfarist
(Parekh) or resourcist (Barry) view of opportunity; or, one must be able sensibly to compare religion with some law-conflicting interests but not others.

Barry, Parekh, Kymlicka, and Sager and Eisgruber have shown that equality can, by way of these other commitments, generate conclusions regarding exemptions. But that their conclusions stand (in some respects) in such direct opposition to one another suggests that the norm of equality itself is not a fruitful starting point for an argument to establish the proper scope of religious exemptions. At the very least, the indeterminacy of equality provides ample reason to look elsewhere for normative frameworks that could generate more determinate conclusions.

In the remainder of this dissertation, I intend to do just that.
CHAPTER TWO:

STABILITY FOR THE RIGHT REASONS

The last chapter identified equality as the predominant basis of impartiality in liberal theory. The next chapter advances an account of impartiality premised on the normative end of stability. It is the project of this chapter to specify what is meant by ‘stability.’ I shall do so by defending a Rawlsian conception of stability – a version of what he calls “stability for the right reasons.” We will be best able to appreciate the merits of this conception by considering it in contrast to an alternative conception advocated by some proponents of “realism” in political theory.

Among the charges levied by realists is that liberals have tended to lose sight of stability as a basic political good. Stability was of central concern to early-modern thinkers such as Machiavelli and Hobbes, but liberalism lost its way when it became overwhelmingly focused on rights-protection and distributive justice. Raymond Geuss, for instance, disdains this approach: “Should political philosophy really be essentially about questions of fairness of distribution of resources? Aren’t security and the control of violence far more important?” Stability is the *sine qua non* of desirable political order, but liberal theory has been captivated by what Bernard Williams called political “moralism,” an orientation which treats political theory as a subset of moral theory – or worse, as merely an “applied” domain of moral theory. Realists such as John Gray, John Horton, and David McCabe have argued that political theory ought to refocus itself on determining the conditions of a workable modus vivendi. Rather than devising unrealizable utopias, political theory ought to be about mitigating disorder and conflict in practical ways.

---

This critique has captured much attention, but, I will argue, it is basically misguided. First, it discounts the attention that liberal philosophers have paid to stability. For instance, the key liberal commitment of toleration or political “neutrality” has been widely understood as a response to the wars over religion that plagued early-modern Europe.\(^3\) Political impartiality with respect to religious identity is valuable (in part) as a means toward order and stability. High liberal paragon John Rawls is a primary target of these realist critics, but their arguments tend to overlook the prominent role granted to stability in Rawls’s work. Recent interpretive work has situated stability as one of Rawls’s primary concerns and suggested that it was concern for stability that motivated his “political turn” between \textit{A Theory of Justice} and \textit{Political Liberalism}\.\(^4\)

Second, the anti-moralism critique advances a normatively flawed ideal of stability, under the rubric of the modus vivendi. As I will argue, modus vivendi stability is vulnerable to persistent challenges posed by individuals’ temptations to defect in pursuit of their individual interests. My objective in this chapter is to show that stability for the right reasons (SFRR) successfully meets those challenges, and for this reason is a preferable normative conception of stability. The argument will proceed in several steps. First (2.1) I will consider the case for modus vivendi, and then (2.2) explain its inherent vulnerability to problems of free-riding and mutual assurance. Then (2.3), I will turn to Rawls’s conception of SFRR, taking some time to


formulate a precise definition of the notion. I show that SFRR is capable of overcoming the free-rider and mutual assurance problems, and then consider three potential objections to SFRR: that it is naïve to the challenges of pluralism (2.4), that it must inevitably fall back on the use of coercion to maintain stability (2.5), and that it is inimical to the actual exercise of politics (2.6). Having rebutted these challenges, I then (2.7) address the question of why stability merits consideration in political theory at all. Then, I situate the conception of SFRR in relation to the ideal/non-ideal distinction in political theory (2.8), and close with some comments suggesting that SFRR may be able to normatively approximate some central requirements of justice.

The aim of this argument is not merely to defend Rawls’s conception of SFRR (though this is indeed part of the aim), but more generally to resist the tendency to neglect the centrality of stability in the liberal tradition. Liberalism has been largely understood – by its critics and proponents alike – as basically concerned with “moralistic” ends such as equality, autonomy, and the protection of rights. As Rawls’s prioritization of SFRR shows, however, even “high liberalism” is not indifferent to this basic political concern. Finally, the defense of SFRR offered here is intended to support my use of that conception of stability in the constructive argument to follow in chapter 3.

Before proceeding, it will be helpful to introduce a distinction. Stability, in the sense in which I intend it here, is not synonymous with stasis or lack of change. I do not define stability in contradistinction to revolution, as anti-liberal thinkers such as Burke or de Maistre might have. These conservatives resisted the 18th and 19th century liberal project of bold social and political innovation, arguing instead that the health of the political body requires gradual, rather than sudden change. To move too quickly, they held, threatens stability. The view of stability

---

that concerns us is otherwise oriented. For Rawls, stability is a feature of “systems,” including those of laws and political institutions (which I will here refer to as ‘regimes’). Stability is a quality that may obtain for a particular “mode and order.” States, on this definition, are not themselves stable or unstable; rather, the regime that governs a state may be either. Stability in our sense is a characteristic of a regime, not a comprehensive description of a political context. In what follows, we will be concerned with regime-stability specifically.

2.1 The Case for a Modus Vivendi

Rawls famously opened *A Theory of Justice* with the claim that “justice is the first virtue of social institutions.” Some realists had to read no further in order to find a point of disagreement with Rawls. Bernard Williams, for example, writes that “I identify the ‘first’ political question in Hobbesian terms as the securing of order, protection, safety, trust, and the conditions of cooperation. It is ‘first’ because solving it is the condition of solving, indeed posing, any others.” Achieving a stable order is a challenge because, as realists tend to see things, politics is inherently marked by the confrontation of divergent, often incompatible, visions and projects. Overcoming or, more likely, containing conflict is the first and most basic political task.

A key premise that informs realists’ assumption of conflict and antagonism is the assumption of robust pluralism sufficient to rule out the likelihood of agreement on substantive

---

6 I recognize that this is a departure from colloquial usage, in which ‘regime’ tends to refer to the administration or collective of persons ruling at a given moment. I adopt the term in the spirit of its usage in ancient authors such as Plato and Aristotle, referring not to a particular set of governing persons, but to the principles and institutions that structure a political society, independent of the persons who occupy any particular offices.


8 *A Theory of Justice*, 3.

9 Williams, *In the Beginning Was the Deed*, 3.

moral matters – including conceptions of justice.\textsuperscript{11} Some of the moral doctrines that citizens hold are bound to conflict fundamentally with others. This means that their desires will likely diverge, as will their conceptions of acceptable political arrangements. To strive for consensus is, according to the realist, a futile endeavor. The “first” task of securing order, then, will need to occur by means other than by getting citizens to agree on basic political questions.

A number of realists have concluded that the goal ought to be the establishment of a modus vivendi arrangement. Though, even those advocates have not been precise in specifying what exactly a modus vivendi \textit{is}. Patrick Neal writes that, in fact, the idea of a modus vivendi remains mostly “the creation of its critics” – such as Rawls – “and it was created for the purpose of allowing them to explain what they are not.”\textsuperscript{12} Straw-man characterizations notwithstanding, it is possible to say something affirmative about what constitutes modus vivendi stability. The basic concept is that a modus vivendi supplies a framework for peaceful coexistence among persons who are bound to remain divided over basic moral commitments. Chandran Kukathas refers to a modus vivendi as reflecting “not epistemic agreement but practical agreement.” Specifically, it is an “agreement to abide by norms which tolerate disagreement.”\textsuperscript{13} Similarly, John Gray describes a modus vivendi as a system of “common institutions in which many forms of life can exist.”\textsuperscript{14} Such arrangements are desirable because, as David McCabe argues, all citizens desire stability, regardless of whatever else they desire.\textsuperscript{15} These descriptions capture one understanding of the \textit{value} of a modus vivendi – namely that it mitigates conflict and permits

\begin{itemize}
\item \textsuperscript{12} Patrick Neal, \textit{Liberalism and Its Discontents} (New York: NYU Press, 1997), 191.
\item \textsuperscript{13} Chandran Kukathas, \textit{The Liberal Archipelago: A Theory of Diversity and Freedom} (New York: Oxford University Press, 2003), 100.
\item \textsuperscript{14} Gray, \textit{Two Faces of Liberalism}, 6.
\item \textsuperscript{15} David McCabe, \textit{Modus Vivendi Liberalism} (New York: Cambridge University Press, 2010), 133.
\end{itemize}
persons to live (within reason) as they like. But this is insufficiently specific; as I will argue below, the same descriptions could be applied with equal accuracy to a regime that meets the higher standard of stability for the right reasons.

A modus vivendi is specifically distinguished by two factors: the way in which it serves the interests of the parties involved, and the reasons for which those parties subscribe to the arrangements.\textsuperscript{16} On the first count, the key to a modus vivendi is that it satisfies parties’ interests only partially. (I take interests here to include both material interests and first-order desires as well as values and normative commitments.) The second condition is that parties to a modus vivendi judge it to represent their best opportunity to have their interests served, given their relative position of power or influence. Gerald Gaus breaks these conditions into four discrete criteria, specifying that some arrangement X is a modus vivendi between parties A and B if and only if:

1. X promotes the interests, values, goals, etc. of both A and B;
2. X gives neither A nor B everything they would like;
3. The distribution of the gains of the compromise (how close X is to A or B's maximum reasonable expectation) crucially depends on the relative power of A and B;
4. For both A and B, the continued conformity by each to X depends on its continued evaluation that X is the best deal it can get, or at least that the effort to get a better deal is not worth the costs.\textsuperscript{17}

A modus vivendi can be represented as the result of a negotiation in which each party’s ability to secure their goals is to some degree limited by the influence of some other party or parties. McCabe suggests that persons will judge modus vivendi arrangements to be a satisfactory “second-best” outcome: unable to secure the mechanisms of political power to advance their particular normative objectives, they will seek instead to secure order and preserve an appreciable degree of latitude to live as they choose. While a modus vivendi arrangement may not give citizens everything they want, McCabe recognizes, “it gives them enough, and in any case, all they can reasonably demand.”  

18 John Horton has described a modus vivendi as a “mix of morality and power,” in the sense that it reconciles diverse parties’ values, desires, and goals with the reality of contest and struggle.  

19 In other words, a realist might say, a modus vivendi represents no small accomplishment. Pace Rawls, there is nothing “mere” about it.  

20 Furthermore, a modus vivendi does not achieve stability by suppressing diversity, but rather provides a stable framework of toleration within which people may pursue different ways of life. For these reasons, a number of realists have concluded, modus vivendi stability ought to be the goal for political societies characterized by pluralism.  

2.2 Two Problems: Free-Riding and Mutual Assurance

In spite of the intuitive, pragmatic appeal of modus vivendi stability, there are certain problems with modus vivendi as a normative ideal. The first thing to note is that some

affirmations of modus vivendi appear to confuse the stability of a regime (or “system”) with that of a political setting in toto. Even if a modus vivendi can be achieved in a given time and place, that is not the same thing as a regime itself having the inherent characteristic of stability. True, a modus vivendi may be the most that can be hoped for in some contexts, but the theoretical concept of regime-stability asks not how to achieve stability here and now, but rather whether a given regime is, in principle, inherently conducive to stability. Modus vivendi may be the most that is possible in many practical settings, but the present investigation requires that we set aside the contingent factors that impede stability in this or that place and time. Still, it is possible to articulate the case for a modus vivendi in terms that do focus on regime-stability. This formulation would suggest that a regime meets the appropriate standard of stability if it capable of winning widespread support as a modus vivendi. So why is this not the appropriate standard?

There are two principal reasons. The first of these is the free-rider problem: citizens will regularly face temptations to “cheat” on their political obligations while continuing to enjoy the benefits of civil order. A stable regime must sustain their desire to comply in the face of such temptation. Second is the problem of mutual assurance: in order for citizens to abide by the terms of social cooperation, they must have confidence that their fellows will do the same. These two challenges are familiar from basic game theory, and they point up the Hobbesian insight that social cooperation poses the challenge of a generalized prisoner’s dilemma. Stability requires both that citizens understand their balance of reasons to tilt in favor of maintaining compliance with the regime, and that they have confidence that others will also choose to comply. Absent these two conditions, citizens will judge it rational to defect, undermining the stability of the regime.

22 Rawls, A Theory of Justice, 295.
23 Ibid., 236.
A modus vivendi can provide some reasons for compliance: citizens recognize that their basic rights are protected and their core interests are served. However, citizens in such a setting will always face temptations to secure their “first choice” political arrangements if circumstances change. A modus vivendi rests on a contingent balance of powers, and rational persons will be ever on the lookout for opportunities to shift the political order more to their liking. This fact motivates the mutual assurance problem, since citizens will always be wary that others will break the terms of cooperation if they can. Also, a modus vivendi does little to address the free-rider problem: citizens who support the regime merely as an acceptable arrangement, but not an inherently desirable one, will be tempted to cheat the system (e.g., non-payment of taxes, draft-dodging, etc.) to their advantage. Under a modus vivendi, citizens’ balance of reasons may be tilted against compliance by two factors: their desire to implement their normatively preferred regime, and their willingness to non-comply with the regime when it serves their interests to do so. A modus vivendi may provide a political framework that citizens can affirm, but this affirmation will always be half-hearted, or at least provisional. Temptations to defect will be more likely to sway them, and their aggregate acts of defection will threaten the stability of the regime.

At this point, the realist might point out that we have ignored the coercive capacities of the political order. Citizens are of course tempted to cheat the system, she will say, but that is why we have penal systems of regulatory oversight and criminal prosecution. In game-theoretic terms, these institutions alter citizens’ payoff tables by imposing sanctions on acts of defection.25 This is, in essence, the Hobbesian solution to the problems of social cooperation.26 Surely, the

---

26 At least on the classic interpretation: see Gauthier, Logic of Leviathan, 86.
realist could say, it is sufficient for stability that a regime resist and manage noncompliance, not that it eliminate all temptations to noncomply.

In practice, penal and corrective institutions are ineliminable features of political order, and can effectively mitigate the destabilizing potential of citizens’ temptations to cheat. The coercive mechanisms of the state are important and (generally) dependable components of orderly societies everywhere. (Also, in anticipation of arguments below, stability for the right reasons is not entirely incompatible with political coercion.) But there are limits to the ability of coercion to secure stability, and the reliance of modus vivendi stability on coercive suppression of noncompliance highlights one of its shortcomings as a normative ideal. For one, stability that does not rely on coercion has lower maintenance costs than stability that requires constant correction by force or the threat thereof. The project of constant deterrence and punishment requires significant resources; much less costly is compliance by intrinsic motivation. (We might call this the Foucauldian insight.) Moreover, when coercion is required to motivate compliance, there exists always the threat of collective non-compliance that could overwhelm the regime’s coercive capacities. The uprisings of the so-called Arab Spring provide recent empirical examples of the limits of coercion as a stability mechanism. The protests – and overwhelming yet ineffectual police response – in Ferguson, MO, in August of 2014 are another. The stability of a regime that takes the Hobbesian approach to motivating support will be costly to maintain, and will be always vulnerable to mass defection.27

---

In short, the stability provided by a modus vivendi is limited in its resilience against citizens’ persistent temptations to noncomply. A modus vivendi is, by definition, susceptible to the free-rider problem, insofar as citizens will be always looking to abandon it in favor of advancing their particular normative agenda. And common knowledge of this fact undermines the conditions for mutual assurance. Recognizing these shortcomings does not require appeal to any “moralistic” considerations or substantive normative concepts. By realists’ own lights, we ought to look for some preferable alternative ideal of stability. My contention is that SFRR is that alternative.

2.3 Stability for the Right Reasons

Most interpreters have treated stability in Rawlsian theory as either a secondary concern or even a distraction from his ostensibly more important concerns – justice, namely. He makes clear that he rejects modus vivendi stability in favor of stability for the right reasons, but he nowhere supplies a comprehensive and precise definition of the latter concept. Additionally, Rawls treats stability only in conjunction with the moral notion of justice. Thus, some work is necessarily to arrive at a concise formulation. For Rawls, a system is stable “whenever departures from it … call into play forces within the system that tend to bring it back to the equilibrium state.” Rawls is concerned most specifically with the stability of conceptions of justice, but, as our formulation of SFRR will show, it is possible to disconnect the notion of stability from Rawls’s focus on justice. Stability itself is formally indifferent to the moral quality of the system (including political principles) in question.

28 E.g., Barry, “John Rawls and the Search for Stability.”
29 Rawls, A Theory of Justice, 400; also see Weithman, Why Political Liberalism?, 44.
Some commentators have read Rawls’s rejection of modus vivendi stability on moral grounds: a modus vivendi is inadequate because it does not guarantee that the principles of justice that are stabilized will be liberal ones.\(^{30}\) This interpretation imputes to Rawls’s entire treatment a moralistic bias: stability is secondary to (a liberal conception of) justice, and a satisfactory conception of stability will take reflect this ordering. In other words, Rawls dismisses modus vivendi stability in favor of SFRR in order to join a non-moral conception of stability with a moral standard of justice.\(^{31}\)

As we will see in section IV, this interpretation underwrites a number of realist objections to the notion of stability for the right reasons. Nevertheless, it relies on a misreading of Rawls’s treatment of stability. As recent exegetical work has shown convincingly, Rawls rejected modus vivendi stability for precisely the reason that a modus vivendi is not sufficiently stable.\(^{32}\) Specifically, Rawls observed, as we did above, that a modus vivendi could not resolve the free-rider and mutual assurance problems. Rawls notes a few problematic features of modus vivendi arrangements. First, they are founded solely on self- or group interests.\(^{33}\) Second, they are the product of political bargaining (actual, not hypothetical), which produces an outcome “compelled by circumstances.”\(^{34}\) And third, consequently, the stability of a modus vivendi arrangement is “contingent on circumstances remaining such as to not upset the fortunate convergence of interests.”\(^{35}\) The paradigmatic example of a modus vivendi is a treaty between two antagonistic states. While the terms of the agreement serve the interests of each state, it will

\(^{30}\) Rossi, “Modus Vivendi, Consensus, and (Realist) Liberal Legitimacy.”

\(^{31}\) Klosko goes so far as to refer to stability for the right reasons as “moral stability.” Klosko, “Rawls’s Argument from Political Stability.” Joshua Cohen similarly says that a Rawlsian well-ordered society is defined by moral consensus. “Moral Pluralism and Political Consensus,” in Philosophy, Politics, Democracy: Selected Essays (Cambridge, MA: Harvard University Press, 2009), 43.

\(^{32}\) Weithman, Why Political Liberalism?, 46–49.

\(^{33}\) Rawls, Political Liberalism, 147.

\(^{34}\) Ibid., 169.

\(^{35}\) Ibid., 171.
be disadvantageous for either to violate it. But this is a tenuous stability: “in general both states are ready to pursue their goals at the expense of the other, and should conditions change they may do so.”  

Rawls does not categorically deny the value of modus vivendi stability. For one, treaties between states are undoubtedly a good, and the fact that a treaty establishes a “mere” modus vivendi does not invalidate it. Moreover, Rawls is awake to the possibility that a modus vivendi could serve as a first step in securing order that might then develop over time into stability for the right reasons; indeed his three-stage genealogy of stability for the right reasons in Political Liberalism begins with a modus vivendi. That a modus vivendi agreement does not secure the most desirable kind of stability is no reason to disparage its appropriate invocation in some circumstances. Nevertheless, the fact remains that modus vivendi stability is essentially, Rawls suggests, a matter of chance. It is grounded in “happenstance and a balance of relative forces,” and cannot be counted on to endure changes in those contingencies.

In sum, Rawls rejected modus vivendi stability not for the reason that it failed to congrue with a standard of justice, but rather for the reason that it fails to secure stability that is resilient to citizens’ temptations to pursue their own interests at the expense of others. Later, this will help us dispatch with some realist objections to the normative ideal of SFRR; for now, it is helpful in specifying what Rawls had in mind by his preferred conception of stability. Whereas modus vivendi stability is vulnerable to defections when citizens reason from within their individual

36 Ibid., 147.
38 Rawls, Political Liberalism, 159.
39 Ibid., 148.
conceptions of the good, SFRR obtains when citizens judge, from the perspective of their conceptions of the good, that it redounds to their good to maintain their commitment to the regime even in the face of temptations to cheat. This occurs when citizens can reflectively endorse the regime from the perspective of their conception of the good. Rawls explicates this conception of endorsement in rather weak terms: citizens’ comprehensive doctrines must be “either congruent with, or supportive of, or else not in conflict with” the principles and values underwriting a political order.  

Put another way, SFRR occurs when citizens regard the regime as 

justified

from the perspective of their particular comprehensive doctrines. Then, citizens will find that their comprehensive doctrines provide support to the political norms of the regime, and that political values “will normally outweigh whatever values oppose them,” defeating the reasons tempting them to cheat on their political obligations.

It merits noting that the concept of justification in liberal theory has largely become associated with a moral commitment to respect for persons that grounds a requirement for coercive laws to be justified to the person(s) coerced. Charles Larmore, for example, says that public justification has “to do with the sort of respect we owe one another in the political realm.” This is characteristic of egalitarian political theories, as we explored in Chapter 1. Here, I am striving to define SFRR apart from substantive commitments like “respect.” Justification, then, should be understood for our purposes as a matter of persons recognizing sufficient reasons to support the regime, not as a matter of respecting the moral status of those persons.

On Rawls’s account, SFRR requires more than just endorsement, though. He also indicates that a regime that is stable for the right reasons “generates its own support,” by which he means the following:

---

40 Ibid., 140, 169.
41 Ibid., 155.
“Its principles should be such that when they are embodied in the basic structure of society men tend to acquire the corresponding sense of justice and develop a desire to act in accordance with its principles. In this case a conception of justice is stable.”

The concept of the “sense of justice” plays an important role in Rawls’s conception of stability. It refers to a desire to comply with the law even when one has reasons to do otherwise. Stability requires that this desire be, as Paul Weithman writes “a highest-order desire that is regulative of [one’s] life.” Even a person who generally desires to comply will face temptations to cheat; a person’s sense of justice is well-developed when her desire to comply with and uphold a regime and its institutions is sufficient to “override disruptive inclinations and … temptations to act unjustly.” The capacity of a sense of justice is innate, but it must be cultivated. Rawls makes clear that stability is a matter of ensuring that persons who “grow up under just basic institutions acquire a sense of justice and a reasoned allegiance to those institutions sufficient to render them stable.” This can occur by either or both of two mechanisms. One is via the formative or educative mechanisms of the state: inculcating loyalty and a desire to uphold political justice. The other is simply by a gradual, organic development of allegiance, where a citizen comes over time to regard the regime as supportive of her good. In either case, the key is that the regime provide the conditions conducive to the development of each citizen’s sense of justice.

43 Rawls, A Theory of Justice, 119.
45 Weithman, Why Political Liberalism?, 53, emphasis in original.
46 Rawls, A Theory of Justice, 398.
47 Rawls, Political Liberalism, 142.
48 Ibid., 163.
We can therefore define SFRR as follows:

A regime is stable for the right reasons if and only if:

(1) Citizens will ordinarily regard it as justified when they reason from within the perspectives of their individual comprehensive doctrines.

and

(2) Citizens who grow up and live under the regime will ordinarily develop a sense of justice sufficient to override temptations to non-comply.

One important feature of SFRR that is embedded in this definition is the quality of inherent stability.\(^{49}\) There are any number of ways in which the second condition – that citizens growing up under a regime will come to support it – could be satisfied, but many of these would be merely accidental. For example, a regime may be stable in an accidental sense when people are too fearful to provoke change, or too lazy. Or, a regime might enjoy accidental stability if a desire to support the regime were to arise in response to a national crisis. These are ways in which persons might (even ordinarily) develop a commitment to their regime, but in each case this support emerges for reasons independent of the normative character of the regime itself. In

\(^{49}\) Rawls refers to “forces within the system that tend to bring it back into this equilibrium state.” *A Theory of Justice*, 400. I adopt the term “inherent” from Weithman, though he counterposes it to “imposed” stability, which appears to neglect the possibility of accidental stability and also confuse the source of stability with its mechanisms. See Weithman, *Why Political Liberalism?*, 45.
contrast, a regime that is inherently stable generates the conditions for its own support by its own means.

That support arises for reasons inherent to the regime does not settle things, however, insofar as a regime has multiple mechanisms for encouraging support from its subjects. For instance, a strong Hobbesian regime generates support, but it does so at the barrel of a gun, ensuring compliance by the threat of force. The coercive instruments of the state are, in the Hobbesian case, inherent to the character of the regime, yet this is not what Rawls has in mind. The stability of a regime or system is predicated of the normative character of the laws, institutions, and principles of the regime, not its penal or coercive instruments.\textsuperscript{50} There are also other ways to secure citizens’ support that do not employ physical coercion or threat thereof, but which nevertheless fall short of SFRR. For instance, consider a regime that cultivated in its citizens an intense fear of attack by other countries. In times of perceived threat, societies have been shown to “rally ‘round the flag” – that is, to increase their support for the ruling regime.\textsuperscript{51} Stability maintained by physical coercion or by psycho-emotional manipulation is tantamount to accidental stability: it does not reflect on the inherent stability of the political principles constituting the regime. Stability for the right reasons, in contrast, is grounded in citizens’ reflective endorsement of the regime. They regard it as justified from the vantage of their own conception of the good, and this confirms their support for it.

Notice that this formulation of SFRR has made no reference to the justness of the regime in question. Indeed, Rawls’s conception of stability is aimed at overcoming the challenges of a generalized prisoner’s dilemma, and does not itself presume that a stable regime must measure

\textsuperscript{50} Weithman says that the stabilizing force must come from the distributive principles that characterize the ruling conception of justice. \textit{Why Political Liberalism?}, 45–46.

up to any particular standard of justice. One of my aims is to show that it is possible to abstract Rawls’s account of stability away from his focus on justice. Instead of a stable conception of justice, we can speak of a stable set of political principles, or a stable regime. The difference may appear merely semantic, but dropping justice helps to clarify that stability is formally indifferent to the moral quality of the political principles in question. At one point in Political Liberalism, Rawls describes stability in terms that evince its independence from moral considerations: “stability is possible when the doctrines making up the consensus are affirmed by society’s politically active citizens and the requirements of justice are not too much in conflict with citizens’ essential interests as formed and encouraged by their social arrangements.” In this characterization, ‘justice’ could be replaced by ‘the regime’ without issue. Nothing else relies on any substantive moral concept.

All that said, in order to appreciate the Rawlsian notion of SFRR in its context, it is helpful to understand how Rawls thought that justice as fairness could achieve stability for the right reasons. As Weithman has shown, Rawls’s “political turn” was motivated by a desire to show the stability of his preferred conception of justice. In his early work, Rawls expressed a confidence that the institutions of a well-ordered state would foster in citizens a sense of justice that would lead them to converge on the ideals embedded in justice as fairness. This is what the Kantian “congruence” argument in Theory purported to show. By his later work, Rawls had recognized that those very same liberal institutions would also encourage a pluralism of comprehensive doctrines, and this pluralism would make it unlikely that citizens would consistently identify the public ideals of justice as fairness with their own private conception of the good. So, in Political Liberalism, Rawls replaced the (partially) comprehensive conception of

---

52 Rawls, Political Liberalism, 134.
53 Weithman, Why Political Liberalism?, 42–67; Freeman, “Congruence and the Good of Justice.”
justice of *Theory* with a “political” (not metaphysical) conception. He then showed that such a conception can become the object of an “overlapping consensus” in which each citizen affirms the political conception from the perspective of her own reasonable comprehensive doctrine. Thus, while the latter rendition recognizes plural reasons for affirming the conception of justice, stability is secured by the fact of each citizen having some such reason.

### 2.4 Pluralism and the Reasonableness of SFRR

The argument in the previous section established that the concept of modus vivendi sets an insufficiently high standard for stability: it would sanction arrangements that are not intrinsically stable, and therefore susceptible to disruption in light of shifting contingent circumstances. SFRR sets a higher bar, rooting stability in universal (or near-universal) endorsement of political principles across a political society. A realist might object that, whatever its conceptual virtues, this standard is thoroughly unrealistic. Specifically, it appears to assume, first, that citizens can reach a robust level of moral agreement in their endorsement of a regime or conception of justice, and, second, that once agreement has been achieved, persons can be counted on to comply with the terms of that accord for reasons other than fear of penalty. Given the fact of pluralism – of which Rawls himself was well aware – the deep consensus involved in SFRR seems to exceed that which is realistically achievable in practice.

Realists, in particular, tend to see politics as inherently marked by the confrontation of divergent, often incompatible, visions and projects. Some of the moral doctrines that citizens hold will conflict fundamentally with others. To strive, then, for broad moral consensus is, according to the realist, a futile endeavor. To place hope in the possibility of stability for the right reasons, it might be argued, is to depend on an unreasonable optimism about the possibility
of consensus. The dictum “ought implies can” is widely accepted as ruling out impossibilities as normative ideals: if SFRR is impossible, then it is a fundamentally invalid political aim. Moreover, a theory that invokes a premise at such odds with observed human psychology is of dubious usefulness. Galston writes that “taking an unattainable standard as the polestar is likely to produce, at best, the frustration of political aims, at worst, destructive distortions of politics.”

This warning of “destructive distortions” points to a deeper worry about overly ambitious political ideals: that they are a mere cloak for more sinister designs. Specifically, one might wonder if the agreement presupposed by SFRR is not cover for an insistence on unanimity – and perhaps a willingness where it does not arise organically to bring it about by force. Chantal Mouffe has defended the claim that all politics is fundamentally hegemonic, including the liberal insistence on consensus as a basis for governance. Given that a political consensus is severely unlikely to obtain, the realist worries that SFRR must fall back on force to impose unanimity. Stuart Hampshire writes that “Domination, the suppression of conflicts by force or by the threat of force, is a great political evil that any citizen may be expected to feel as evil...” Critics of liberalism have long charged that while liberals espouse freedom and choice, they are just as interested in imposing their preferred way of life as those who they would be quick to decry as anti-liberal. In light of the profound differences which trouble any hope of agreement, can stability for the right reasons come about through any means but coercion of those who resist consensus?

54 Galston, “Realism in Political Theory,” 395.
This might sound like too strong a charge against Rawls. How can the theorist of “the fact of reasonable pluralism”\textsuperscript{58} be implicated in the liberal hegemony this criticism alleges? Actually, the criterion of “reasonableness” points up the issue. A number of commentators – not all within the realist camp – have criticized Rawls for his too-willing exclusion of “unreasonable” comprehensive doctrines from the enterprise of public justification.\textsuperscript{59} Rawls describes a reasonable person as one who is “ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so,”\textsuperscript{60} but more narrowly supposes that “a reasonable comprehensive doctrine does not reject the essentials of a democratic regime.”\textsuperscript{61} Those who contest these democratic essentials are not merely disagreeable, but \textit{unreasonable}. Rawls states that the presence of such persons introduces “the practical task of containing them – \textit{like war and disease} – so that they do not overturn political justice.”\textsuperscript{62} Read with a realist lens, this commitment to “containing” dissent might be seen to evince an easy comfort with coercion.\textsuperscript{63} Bonnie Honig finds in Rawls’s work a tendency to criminalize and punish those who do not affirm the reigning conception of justice, treating these sources of disorder as “outside agitators.”\textsuperscript{64} In order to maintain his ideal of a political order that is self-sustaining and self-reinforcing, Rawls is forced to resort to blunt coercion to deal with those who remain outside the consensus. This makes stability for the right reasons seem not only naïve, but even dangerous.

\textsuperscript{58} Rawls, \textit{Political Liberalism}, xix.
\textsuperscript{60} Rawls, \textit{Political Liberalism}, 49.
\textsuperscript{61} Ibid., xviii.
\textsuperscript{62} Ibid., 64n19 (emphasis added).
There is no denying that SFRR sets an ambitious standard. Formulating a regime that is capable of achieving SFRR is no easy task, and achieving the support it demands in actual practice even less so. To claim that SFRR is, in spite of these difficulties, an appropriate normative ideal for politics reflects a definite optimism. But ambition and optimism are not the same as naivete. The question is not whether a normative ideal can be easily or immediately actualized, but whether it represents a realistically possible goal. The realists are correct to point up the challenges facing SFRR, but a closer look at these challenges shows that there is no reason given by either abstract reasoning or empirical observation to believe that they cannot be overcome.

Rawls’s later work, in particular, is fully cognizant of the inevitable development of incompatible comprehensive doctrines. In *Political Liberalism*, he writes that “Under the political and social conditions secured by the basic rights and liberties of free institutions, a diversity of conflicting and irreconcilable – and what’s more, reasonable – comprehensive doctrines will come about and persist if such diversity does not already obtain.” This is a product of both liberal institutions and the “burdens of judgment.” In assuming the inevitability of pluralism and disagreement, Rawls and the realists are actually on the same page. But from the fact of pluralism it does not follow that citizens affirming diverse moral doctrines cannot come to support common political principles. In fact, such convergence is perfectly ordinary: A opposes unnecessary killing because it violates the categorical imperative, B because it violates the biblical commandment “Thou shalt not kill.” From two fundamentally incompatible sources we derive the same moral conclusion. As a matter of logic, it simply does not follow from the fact of moral pluralism that moral convergence on political principles is impossible.

---

The possibility of such convergence is the basic insight behind Rawls’s conception of the overlapping consensus. An overlapping consensus occurs when citizens who hold diverse and incompatible comprehensive doctrines come to support a single political conception of justice. Rawls held, though, that an overlapping consensus requires not only endorsement of common principles but also that those principles take the form of a “freestanding” conception of justice that does not depend on any one comprehensive doctrine for its justification. Charles Larmore interprets this to mean that citizens’ “reasons for embracing the principles must not spring simply from their different perspectives but must also draw upon a common point of view.” This suggests that agreement on principles is insufficient; citizens must also agree on the reasons underlying the principles. This would seem to ask too much in conditions of pluralism: A and B may agree on certain norms, but they are unlikely to find a moral “common point of view” that they can both embrace.

A number of authors have contested the insistence on a freestanding conception of justice as an element of an overlapping consensus. Gerald Gaus and Kevin Vallier, in particular, have convincingly shown that what is actually required is not a robust consensus, but rather modest convergence on political principles. The stability of a conception of justice that can win endorsement from the perspective of various citizens’ incompatible comprehensive doctrines is not weakened by virtue of the fact that each citizen’s reason for endorsement is different. Indeed,

66 Ibid., 165–66.
Vallier argues that convergence from diverse perspectives “respects reasonable pluralism by helping citizens recognize that others often have distinct reasons to endorse a proposal,” which supports stability by providing mutual assurance.⁶⁹ Contra Rawls, consensus on a freestanding conception of justice is a standard unnecessarily high for stability.

Rawls suggests that agreement at only the level of principles (which he calls a “constitutional consensus”) is faulty because it “lacks the conceptual resources to guide how the constitution should be amended and interpreted.”⁷⁰ But even if there is disagreement over amendment and interpretation, these need not be fatal to stability. First, disagreement over whether or how to change standing laws and institutions does not trouble commitment to the status quo. Disagreement might impede amendments to the constitutional order, but it would not necessarily produce defections from it. Second, it is not clear that even a “common point of view” would avoid the issue of interpretive conflict. Interpretation by definition implies ambiguity, ambivalence, and disagreement. Even where citizens affirmed a freestanding conception of justice, replete with guiding ideals such as the conception of citizens as free and equal, citizens might reasonably disagree on the proper meaning, scope, or weight of some component principle. Conflict of this sort is not a problem, but rather an intrinsic element of interpretation. Whether this conflict occurs within the scope of shared “conceptual resources” or across competing comprehensive doctrines that nevertheless support a common set of political principles, it poses no necessary threat to stability. Citizens may disagree about what exactly a principle means while maintaining their shared commitment to supporting and complying with it while they work toward interpretive agreement.

⁷⁰ Rawls, Political Liberalism, 165.
None of this is to say that consensus is *incompatible* with the ideal of SFRR. Rather, consensus at the level of reasons underlying political principles is *unnecessary*. Convergence on support for those principles is sufficient to secure the conditions of SFRR. What these share in common is that, in the cases of both consensus and convergence, citizens’ support for political principles is a matter of outright moral endorsement. Above and beyond these, there are additional modes of support, not rooted in outright moral endorsement, that are also sufficient to secure stability for the right reasons.\(^71\) These include at least the following: normative strategy, an independent commitment to reciprocity, and enlightened self-interest.

The first of these marks a distinction from the kind of strategic endorsement that is characteristic of a modus vivendi. A modus vivendi involves what we might call interest-based strategy, as it involves citizens embracing political arrangements on the basis of how well they serve their basic interests. They set aside their primary goal – the advancement of a normative project – in favor of more basic goods such as liberty.\(^72\) Normative strategy, in contrast, flows directly from a desire to advance one’s fundamental commitment to a comprehensive doctrine, albeit modulated by recognition of practical/contingent limitations. A citizen may judge that a given political principle, while it may not accord perfectly with her moral beliefs, nevertheless offers a practically accessible means to advance her moral doctrine. For instance, a moral egalitarian might judge that strict enforcement of absolute equality in all realms would likely generate resentment and actually undermine some citizens’ support for egalitarian aims. On the basis of this concern, the egalitarian could support a regime admitting of some inequality on grounds that doing so is the best way to secure equality’s ultimate advance. The principles

---


\(^{72}\) See McCabe, *Modus Vivendi Liberalism*, 133.
governing this regime would fall short of outright endorsement by her moral doctrine, but the
citizen’s affirmation of these principles is not “merely” strategic in the modus vivendi sense – it
is part of a normative strategy. The legitimacy of normative strategy is supported by the fact that
translating a moral view into viable political commitments always involves some consideration
of contingent facts as they bear on the application of moral norms. The distinction between
outright moral endorsement and normative strategy, therefore, is not a hard and fast one, and
there is no reason to think that only the former is a potential component of SFRR.

Second, a citizen’s moral doctrine might affirm not a specific political principle, but
rather the moral worthiness of social cooperation itself, and maintain support for the regime as a
matter of reciprocity, rather than on the regime’s particular merits. Such a person might hold a
moral view that entails a commitment to identifying principles that others can also agree to live
by, and then maintaining one’s promise to live by them (whatever their content). This is close to
what Rawls has in mind by the desideratum of “reasonableness” as a characteristic of some
comprehensive doctrines. A commitment to reciprocity is formally neutral in regard to the
content of the principles; it affirms the good of cooperation under whatever principles can win
support from others.

Third, citizens might sometimes affirm political principles on the basis of considerations
that are strictly prudential – that is, having no basis in a comprehensive moral doctrine – but that
are nevertheless profound in such a manner as to motivate a resilient commitment of support.
Ordinary calculation of self-interest cannot alone provide such a basis, for it would be vulnerable
to the free-rider problem: if I support the regime for the reason that it serves my interests, then I
will be tempted by the possibility of advancing my interests even further by cheating. But a more
sophisticated conception of self-interest, such as Tocqueville’s doctrine of “self-interest properly

73 Rawls, Political Liberalism, 49.
understood,” could do the trick. A long-term, broad-scope conception of self-interest might lead a citizen to conclude that her critical interests are best served by maintaining a commitment to compliance. Of course, not all citizens will be capable of reasoning their interests on these terms, but some will, and for these persons it could generate the commitment necessary to overcome the temptation to free-ride when it arises.

We have therefore identified at least five modes of support that a regime might enjoy. The first two – consensus and convergence – rely on citizens’ moral endorsement, while the other three – normative strategy, reciprocity, and enlightened self-interest rest on other factors. Only one of these, consensus, presumes robust moral agreement. Insofar as this kind of agreement occurs, then it surely contributes to the stability of a regime. But SFRR neither assumes nor requires that moral consensus will arise. All SFRR requires is that citizens converge in support of a common regime. This support may be grounded in any of the sources described above – and there are surely other possibilities. In many cases, a citizen’s commitment to compliance will be over-determined: they will have more than one sufficient reason comply. The basic point is that realists’ justified skepticism about the possibility of moral consensus does not rule out SFRR as a reasonable aim, as there are plenty of alternative motivational sources available to support citizen’s convergence on political principles. SFRR is therefore entirely compatible with recognition of robust pluralism. The standard it sets is undeniably ambitious, but not so much as to be unrealistic or naïve.

---

74 *Democracy in America*, II.i.8: “How the Americans Combat Individualism by the Doctrine of Self-Interest Properly Understood.”
2.5 Stability and Coercion

What, though, about the truly intransigent citizen who refuses to compromise, to yield, or to endorse anything less than full-bore enshrinement of his particular normative beliefs in laws and political institutions? This might include the cases as disparate as the egotist who cheats relentlessly on his taxes and the Christian county clerk who refuses to sign a marriage license for a same-sex couple. Even if it is reasonable to hope for broad convergence around support for a set of political principles, the profound pluralism of liberal societies is bound to include at least some who will fail to join the convergence. Does the fact that persons such as this are a real and (probably) ineliminable feature of social life mean that any robust conception of stability must commit to the forced assimilation of unreasonable persons? Must stability for the right reasons fall back on the use of coercion to “contain” them, as Rawls suggests?

One has no option but to concede that stability will, in practice, require at least some coercion of some unreasonable persons. This admission, though, does not fatally undermine the ideal of stability for the right reasons. SFRR, understood as explicated above, does not imply that each and every citizen will successfully resist each and every temptation to non-comply with the regime. Rather, their sense of justice may be sufficiently robust as to ordinarily override these temptations. SFRR does not rule out the possibility that citizens will, in some instances, succumb to those temptations. Otherwise it would imply that stability requires universal perfect compliance – surely too high a standard to be realistic. I would suggest that there are three kinds of cases where coercion would be warranted to penalize non-compliance as a means of correcting for specific problems of political order. Those problems, which I will detail

---


76 Rawls himself recognizes this. A Theory of Justice, 504.
momentarily, are deficient rationality, antisocial personality, and imperfect information. The key is that in each of these cases, coercion is permitted as a mechanism for securing stability in non-ideal conditions. The coercion does not function as a substitute for the capacity of the regime to generate its own support, but rather as a corrective to threats to stability that do not arise from the character of the regime itself.

In the first type of case, coercion serves to deter and penalize citizens who, while being in general committed to supporting the regime, may fail to translate their general commitment into perfectly compliant conduct. That a person may sincerely believe that the regime merits support on the basis of his comprehensive doctrine does not mean that he will act in uniformly compliant ways. While non-compliance may reflect deliberate and rational decisions to cheat on obligations, it may also reflect a failure to derive compliant maxims of action from sincerely-held beliefs that should lead to compliance. In cases like this, the acts of non-compliance actually violate the citizen’s own beliefs. It is not supposed, then, that they reflect some deficiency of the regime, but rather a deficiency in the citizen’s reasoning faculty. Coercion is therefore consistent with SFRR as a corrective for this deficiency. This is likely to be the most common role for coercion in a stable regime.\(^77\)

The second type of case concerns those atypical persons who are simply disposed consistently to act in socially disruptive ways. These persons may need, literally, to be contained, as their materially damaging actions pose a threat to social order. But their threat to stability is not contagious – psychopathic serial killers, virulent Klan members, and violent religious fanatics do not normally tempt others to follow suit. But their actions may be materially

\(^77\) It is hard to delineate precisely which rational deficiencies are uncharacteristic and which are perfectly ordinary – and therefore must be taken into account by a regime that is to be stable. For our purposes, we only need recognize that there are some instances of non-compliance that do not reflect any general diminution of support, but are rather products of particular circumstances or localized factors.
damaging to others, and if they go unpunished, they weaken citizens’ confidence that the political order effectively protects their interests. Coercion in these cases is analogous to rules governing hazardous materials or firearm use, which serve to protect others from harm. Coercion of antisocial personalities does not aim to motivate the antisocial person to support the regime, but rather to protect the interests of other citizens so that they will maintain their support.

The third class of cases where coercion may be consistent with SFRR involves non-compliant acts that themselves are not materially detrimental to stability. For example, a subway turnstile jumper poses no real threat to the stability of the transit system – the revenue lost by a single unpaid entry is miniscule. But even one turnstile jumper can have a detrimental effect on stability by virtue of how other citizens perceive and interpret that individual violation. A citizen may observe a jumper and infer a widespread problem (as opposed to an isolated instance of rule-breaking) in recognition of which the observer feels like a sucker – paying his entry when so many others enter for free. Or, seeing an unpunished subway jumper, a citizen might infer that the system must really not depend on fares as an important revenue source. Either of these inferences may motivate the observing citizen to avoid paying his own fare. Both are possible because of the citizen’s incomplete and imperfect information, and suggest how coercive penalties can mitigate the indirect damages to stability that even materially inconsequential acts of non-compliance can generate. Penalties prevent citizens from being tempted to non-comply on the basis of imperfect information.\(^{78}\)

The people that Rawls called “unreasonable” might thus be handled under any of these rubrics. Coercion may be compatible with SFRR when it addresses the problems of deficient reason, antisocial personality, or incomplete reason. But this is not a warrant for unlimited coercion. Notice two assumptions embedded in the turnstile-jumping example: that the number

of jumpers is relatively small, and that the observing citizen is not motivated to cheat by his simple desire to save a buck. In practice, the main problems reflected in even small acts of non-compliance are the basic challenges of free-riding and mutual assurance: even reasonable citizens who see others break the rules may be tempted to similarly evade their obligations. SFRR does not admit of coercion to overcome these challenges; this is the Hobbesian solution. A regime that is stable for the right reasons relies instead on citizens’ reflective endorsement to motivate support. I have suggested that coercion can be consistent with SFRR to correct for non-ideal conditions regarding rationality, personality, and information – deficiencies that are not intrinsic to the character of the regime itself. But for the basic challenge of motivating citizens to pay their subway fare, or pay their taxes, or abide by the law in general, coercion is not a permissible solution.

This points up the ambitiousness of SFRR: while some coercion is justified, the ideal requires that citizens generally comply with their legal obligations for reasons other than threat of punishment. Designing a regime capable of winning this kind of support is no simple task (though Rawls thinks he has done it with justice as fairness). What I have claimed is that such a regime is possible. The fact that, in most actual situations, we rely on Hobbesian penalties in order to maintain compliance does not count against SFRR. It merely reflects how far we are from achieving it. But stability is not a binary condition. We can say that a regime is more or less stable depending on the extent to which it relies on coercion to secure support. Also, the ideal of SFRR remains a useful orienting goal even when it is not perfectly realized. As Rawls describes in the three-stage account Political Liberalism, stability for the right reasons may be the end result of a political evolution that involves persons developing confidence in liberal institutions
and, consequently, acquiring more reasonable dispositions over time.\textsuperscript{79} Thus, even in places where SFRR is not an immediately realizable goal, it is a reasonable aim for the long term.

2.6 The Politics of Stability

One of the signature commitments of realism in political theory is a conviction that politics is fundamentally a playing field for power. Its corresponding critique is that “high liberalism represents a desire to evade, displace, or escape from politics.”\textsuperscript{80} Many realists hold that politics is not some domain of reasoned equanimity and eager collaboration, but is rather, as Marc Stears glosses it, “inherently a process or a practice that involves some form of domination.”\textsuperscript{81} On this picture of politics, stability will be achieved by the successful imposition of power, or by the achievement of a balance of power – not by somehow eliciting a pacific desire on the part of citizens to comply with the law. “Agonist” theories of politics, in particular, prize the contestatory features of politics. Chantal Mouffe, for example, has rejected the possibility of moving “beyond hegemony” in politics.\textsuperscript{82} The realist movement aims both to unmask the hidden operation of power in political society, and to promote a form of political order that confronts honestly the fact of contest.

From this vantage, stability for the right reasons might appear to envision a profoundly anti-political equilibrium, a sterile consensus instead of a fertile domain of dispute. John Gray writes that “Rawls’s is a liberalism which has been politically emasculated, in which nothing of importance is left to political decision, and in which political life itself has been substantially

\textsuperscript{79} Rawls, \textit{Political Liberalism}, 163.
\textsuperscript{80} Galston, “Realism in Political Theory,” 386. See also Glen Newey, \textit{After Politics: The Rejection of Politics in Contemporary Liberal Philosophy} (Palgrave, 2001), 22–30; Honig, \textit{Political Theory and the Displacement of Politics}.
\textsuperscript{82} Mouffe, \textit{On the Political}, 3.
evacuated of content.” Even if SFRR is a reasonably realistic ambition, the realist could nevertheless object that it deceptively obscures the role of power in politics, and therefore merits rejection as an *inappropriate* standard for the political domain.

This rejection, however, rests on a misunderstanding of stability for the right reasons. Namely, it assumes that SFRR requires deep moral consensus. It is true that when everyone agrees, there is nothing left to argue about. But, as I have acknowledged above, such consensus is an impossibility in liberal societies. SFRR does not rest on hope of consensus, but rather aims at convergence around support for a set of political principles. This convergence does not presume a monochrome world of universal agreement, but rather an intersection of plural motivations.

This is entirely compatible with robust contest and struggle. Rawls, for one, intended no suppression of disagreement, insisting that agreement at the level of the overlapping consensus concerns only things like “constitutional essentials,” or the “basic structure” of a political society, not the full domain of the political. The better part of social life remains subject to the ordinary back-and-forth of political contestation. Rawls is clear that the “background culture” of a well-ordered society “contains comprehensive doctrines of all kinds: these are taught, explained, debated one against another, and argued about, indefinitely without end, as long as society has vitality and spirit.” Power will inevitably be at work in these dynamics. As a feature of ordinary democratic discourse, this is normal and likely unavoidable. The realist hermeneutic is invaluable for assessing stability, however. A convergence that is supported not by reflective endorsement but rather by coercion is not inherently stable, and therefore falls short of the ideal

---

85 Ibid., 383.
of SFRR. Attention to the authenticity of convergence is an important contribution of the realist mode of political analysis.

There are limits to the desirability of contest and struggle, however. Even those who affirm an agonistic conception of politics must concede the desirability of some stable – if minimal – order to “channel” conflict and disagreement. Stability for the right reasons does not define the scope of the set of political principles to which it applies; it is a formal idea. In some cases, a wide set of principles covering a great deal of public life may achieve SFRR. In others it may be just a narrow set of principles that set basic parameters for the practice of politics. Rawls would seek to enshrine a thicker set of political norms in this stable order than hardcore agonists would likely prefer, but the difference is one of degree rather than of kind. SFRR does not presume anything about the scope of the subject of stability, and is perfectly compatible with the maintenance of a robust sphere of political conflict and dispute.

It is necessary to concede, though, that any idea of stability involves removing some components of political order from the field of play. Stability means “settling” some issues in such a manner that they may endure, whatever other political arrangements may change. If an agonist insists that this is an undesirable ideal, then we may have reached an insuperable impasse. If “true” politics must involve constant struggle over the entire domain of public order, then an ideal of stability has no place. But this image of pure contestation is a thoroughly frightening possibility – one that political theorists have rejected time and again. Hobbes is the most obvious example, describing the political order as a conventional arrangement that, while not necessarily pretty, has the basic virtue of saving us from the horrors of the state of nature. Machiavelli, too, while more attentive to the inevitability of flux and danger in political life,

---

86 Galston, “Realism in Political Theory,” 397.
87 Rawls, Political Liberalism, 151–52.
reminds the Prince repeatedly of the evils of disorder. More recently, Judith Shklar’s “liberalism of fear” has drawn attention to the core function of political order as a bulwark against cruelty. An ideology that romanticizes political struggle absolutely is wrong-headed, and appears to reflect a false nostalgia born of a failure to appreciate the distinct advantages to living in a context of political stability. Even if Nietzsche’s critique of the domestication wrought upon humans by modernity is descriptively persuasive, an uncritical return to the wild is no suitable remedy.

2.7 Why Stability in the First Place?

Thus far, I have defended the end of stability for the right reasons as the conception of stability most suited to be adopted as a normative ideal. After showing its superiority to modus vivendi stability, I defended it against three critiques: the first alleging that SFRR is unrealistically ambitious, the second that it must fall back on the very coercion which it avows to eschew, and the third that SFRR necessarily elides genuine politics from its conception of political order. I have tried to show that these objections rest on a misunderstanding of SFRR as involving a moralized conception of consensus. Rather, I have argued, SFRR is a realistic notion of convergence designed to overcome the persistent threats to stability posed by individuals’ temptations to defect. Seen in this light, SFRR is not merely an acceptable version of liberal stability. It is the conception of stability that realists themselves ought to endorse.

I now want to consider a challenge from a different direction, namely one that rejects any normative ideal of stability. Where a realist might accuse SFRR of too much high-minded

idealism, the critique I wish to take up now dismisses concern for the real-world stability of a theory as an unnecessary capitulation to contingent constraints.

Stability is a sub-concern of what is often referred to as “feasibility,” which focuses on the real-world possibility of actualizing a theoretical political conception.\(^9\) To care about stability is to be concerned not with the rightness or justice of a set of political principles, but with whether they can be actualized and endure – i.e., whether they represent a feasible political program. Juha Raikka suggests that “desirability” and “feasibility” approaches to political theory focus on distinct questions: the former on whether a normative proposal is in itself choiceworthy (to employ Aristotle’s term), the latter on whether it sits within the domain of the possible.\(^90\) Stability is a component of the feasibility approach, in that stability is a necessary (but not sufficient) condition for feasibility: if a set of political principles is not stable, then it is not feasible.\(^91\) Rawls, for one, is clear that he intends his account of political liberalism to concern itself with “practical political possibilities.”\(^92\) As I mentioned above, “ought implies can” has often been taken to mean that any true normative political principle must be actualizable.\(^93\) Yet, some political theorists have pushed back against the feasibility approach on grounds that it inappropriately constrains the activity of philosophical theorizing about politics. In short, this

---


\(^91\) Ibid., 32n16; Laegaard, “Feasibility and Stability in Normative Political Philosophy,” 406; Gilabert and Lawford-Smith, “Political Feasibility,” 811. It might be objected that a political program that lasted for some appreciable amount of time (a decade, say) and then collapsed might still qualify as “feasible,” undermining the claim that stability is necessary for feasibility. Since we are here concerned with inherent stability, I would interpret an occurrence like this to indicate that the program enjoyed a meaningful, but limited, degree of stability. From this it would follow that the feasibility of the program is also limited. A program that is prone to collapsing, even if it can endure under some conditions, could reasonably be described as a non-feasible conception of justice.


\(^93\) E.g., Mason, “Just Constraints,” 256.
objection holds that feasibility is irrelevant – or, at best, of only secondary relevance – to the
determination of normatively worthy political arrangements. Whereas the realist found SFRR
insensitive to the facts of human social life, this latter critique objects to a focus on stability as
being too sensitive to contingent facts.\(^94\)

David Estlund, for instance, rejects what he sees as “utopophobia” in political
philosophy, noting that moral theorists do not impose the same kinds of constraints on their
philosophizing: “Moral philosophers know that people are likely to lie more than they morally
should, but this doesn’t move many theorists to revise their views about when lying is wrong.”\(^95\)
In contrast, political philosophers tend to reflect at least a minimal realism in their work,
eschewing normative requirements that they judge as impracticable or unlikely to obtain. Estlund
finds this disposition unfounded, and insists that “unrealistic” theorizing is important to political
philosophy.\(^96\)

Estlund draws an analogy to a mundane experience of aspiring: “Suppose we are hiking,
and we spy a beautiful spot some miles off... Alas, it is not yet clear whether we can get there, so
we might try to contain our excitement.” The feasibilist approach would counsel us to “Be
realistic. Things are fine where we are, so we could just conclude that the new spot is not really
worth considering.” But Estlund suggests that this is a short-sighted response. “[I]f we are not
sure that it is impossible, then even if we are unlikely to get there, it could be worth thinking
about how we might. … After all, the place is beautiful, and for all we can tell getting there is

\(^{95}\) David M. Estlund, Democratic Authority: A Philosophical Framework (Princeton, N.J: Princeton University
Press, 2008), 12.
\(^{96}\) Ibid., 263–70; “Human Nature and the Limits (If Any) of Political Philosophy,” Philosophy & Public Affairs 39
not impossible.”  That we can be sure of neither the possibility nor the means of achieving some desirable end need not imply that we should shun that end in our thinking.  

Estlund’s rejection of “utopophobia” contains a couple of related critiques. The first of these worries that an emphasis on feasibility leads to complacency in political theorizing. Our judgments of what is possible and what is impossible are inevitably colored by contingent context and experiences. The more remote from current practice a proposed norm is, the more likely we are to judge it an impossibility. If we reject out of hand any political principles that appear impossible to implement, we may unwittingly contribute to the perpetuation of the status quo. This is presumably what Rousseau had in mind when he wrote, “‘Propose what can be done,’ they never stop repeating to me. It is as if I were told, ‘Propose doing what is done.’” A relentless focus on feasibility can blunt the critical edge of political philosophy.

The second component of Estlund’s critique accuses adherents of the feasibility approach of conflating the categories of possibility and likelihood. There is a fundamental difference between standards that cannot be met and standards that are unlikely to be met. A theory that requires of people something that is definitively impossible is indeed faulty, as it rests on a false premise. In contrast, a theory whose requirements are lofty, but possible, has no defect. Compliance may be difficult or unlikely, but this is not the same as it being impossible. The feasibility approach all too readily writes off as inappropriate not only impossible standards but

---

97 Estlund, Democratic Authority, 269–70.
100 Rousseau, Emile, preface, par. 5, quoted in Estlund, “Utopophobia,” 115.
unlikely ones as well, but this is a mistake. As Estlund writes, “it is not the case that ought implies reasonably likely.”

These two critiques suggest that the feasibility approach is the wrong orientation to adopt toward political theorizing. Normative theorists ought to focus on the desirability of political principles and, so long as they are not categorically impossible, leave the questions of whether and how we might travel from here to there as a secondary matter. To reject the feasibility approach, though, would suggest that to situate stability centrally as a regulative end in our political theorizing is to impose an unnecessary constraint on normative analysis of politics. Why limit ourselves to those political principles that we recognize can be stable over time? Why not identify those political principles which would truly instantiate justice and only then, as a secondary concern, wonder how they might be rendered stable?

Paul Weithman has suggested that Rawls’s personal motivation in seeking a conception of justice that could be stable for the right reasons derived from deep-seated religious questions that prodded Rawls for much of his life. The horrific events of the mid-twentieth century – the Holocaust and World War II, in particular – deeply troubled Rawls. He understood these as posing a fundamentally theodical question (one which he traced to Kant): “whether it is worthwhile for human beings to live on the earth?” If the answer is yes, for Rawls, then “a reasonably just society must be possible,” and the project of his life’s work was to demonstrate this possibility. To show that a just society is possible – and that it could be sustained by means other than by the forceful bending of human inclination – was Rawls’s answer to the sorrowful historical evidence pointing in the direction of humanity’s depravity. Valuing stability

---

101 Ibid., 116.
102 Weithman, Why Political Liberalism?, 368. (citing Rawls, Political Liberalism, lxii).
103 This quote comes from an unpublished draft of the Introduction to the Paperback Edition of Political Liberalism, cited on Weithman, Why Political Liberalism?, 368.
for the right reasons in political theory reflects no less than a commitment to answering the theodical challenge in an affirmation of humanity’s goodness.\textsuperscript{104}

This response is moving, but it says more than is necessary to defend a concern for stability in normative theorizing. A simpler answer points instead to the strong connection between the stability of a political principle and the desirability of that principle. Stability concerns the likelihood of a political arrangement enduring once implemented. While this might seem like a quintessentially feasibilist concern, I would contend that if a principle is unlikely to be sustained over time (for reasons of human psychology, for example), this in fact counts against the desirability of the principle.\textsuperscript{105} To affirm a political principle as desirable is to affirm that a world in which that principle was instantiated in institutions is a desirable world. But to instantiate in institutions an inherently unstable principle is to invite the failure of those institutions.\textsuperscript{106} To prefer normative principles that are bound to fail to endure is tantamount to preferring chaos – and the consequent harms that are bound to follow. As mentioned above, I take the insights of Hobbes and Machiavelli to definitively rule out disorder as a desirable state of affairs. A principle that cannot be stable is not desirable.

This is true of stability in a way that it isn’t for some other components of feasibility. Consider the accessibility of a normative principle, for example: the possibility of instantiating that principle in a world roughly like ours – getting from here to there, as it were.\textsuperscript{107} As Estlund

\textsuperscript{104} This reading addresses Raymond Geuss’s reasonable perplexion at Rawls’s claim that his attempt to articulate a theory of justice was inspired by the horrors he witnessed in the second World War. See Geuss, \textit{Outside Ethics}, 29–31.
\textsuperscript{105} Rawls, \textit{A Theory of Justice}, 496–99.
\textsuperscript{106} Gilabert and Lawford-Smith argue that stability is actually a necessary condition of accessibility: “we can see that stability is in fact part of the package of feasibility considerations by noticing that getting ‘there’, if we stay there for only a short while, does not really look like a case of ‘getting there’ at all.” Gilabert and Lawford-Smith, “Political Feasibility,” 813.
rightly notes, the fact that contingent features of our current context render it (apparently) impossible to implement a principle does not imply anything about the desirability of the principle. Reflection on this principle can still illuminate important features of normative political categories, and it can provide a normative “target” to strive for, even if perfect actualization is not possible. The accessibility of a political norm is an accidental characteristic, dependent on the contingencies and path-dependencies of actual political contexts. The stability of a norm, in contrast, is of intrinsic significance, insofar as it tells us whether the norm might manifest in an enduring order, or whether it is bound to disintegrate upon arrival.

2.8 Situating Stability in Ideal and Non-Ideal Theory

The question of desirability vs. feasibility can be easily conflated with another distinction between genres of political theory: ideal vs. non-ideal. The question of whether a concept of stability for the right reasons belongs to ideal or non-ideal theory is closely related to the feasibility/desirability question: how far does this conception of stability descend into the contingent mess of actual political constraints? But the ideal/non-ideal distinction is distinct from the desirability/feasibility issue. Here I want to offer some suggestions on how to situate the concept of stability for the right reasons – and my project more broadly – in relation to this latter distinction. I should note that this section does not carry any significant weight for the main argument of this chapter, but goes some way to situating the argument of this dissertation in a theoretical context.

---

110 It is also distinct from the liberalism-realism divide. See Matt Sleat, “Realism, Liberalism and Non-Ideal Theory Or, Are There Two Ways to Do Realistic Political Theory?,” Political Studies, 2014.
The distinction between ideal and non-ideal theory is as old as political theory itself (see Plato’s “just city in speech”), but the contemporary articulation comes from Rawls, who characterized his own theory as an exercise in the former category. The distinction is basically intuitive: ideal theory involves constructing “a conception of a just society that we are to achieve if we can,” while nonideal theory “asks how this long-term goal might be achieved, or worked toward, usually in gradual steps.” Ideal theory strives to describe a society that is perfectly just; non-ideal theory wrestles with how to achieve justice from unjust circumstances. Pursuit of the aims of ideal theory requires that the theorist adopt some idealizing assumptions. Rawls names two in particular: strict compliance with the principles of justice, and favorable circumstances, including historical, social, and economic conditions. Adopting these assumptions allows the theorist to evaluate the merits of alternative political principles directly, controlling, so to speak, for the variables of compliance and circumstance. Thus, the assumptions involved in ideal theory support the desirability aim: they help the theorist identify the political principles that are intrinsically best, apart from their likely efficacy if applied to some given real (i.e. non-ideal) context. Non-ideal theory does without these idealizing assumptions, and strives for comparative assessments of imperfectly just proposals or pragmatic routes toward the achievement of justice.

Given these characterizations, it would seem that a project that begins from stability, as mine does, must fall into the camp of non-ideal theory. Non-ideal seems an obvious descriptor

---

111 Rawls, A Theory of Justice, 216.
112 John Rawls, The Law of Peoples (Cambridge, MA: Harvard University Press, 1999), 89. For a thorough reconstruction of Rawls’s conception of the ideal/non-ideal distinction, see Simmons, “Ideal and Non-Ideal Theory.”
115 Simmons, “Ideal and Non-Ideal Theory,” 8–9, 15–16.
116 I will leave aside the question of whether one or the other of these two aims most properly characterizes non-ideal theory. See Simmons, “Ideal and Non-Ideal Theory”; Robeyns, “Ideal Theory in Theory and Practice.”
for a discussion of how to keep a regime from collapsing into chaos. But, in actuality, it is not necessary to exclude an investigation of stability from the purview of ideal theory.

For one, I intend to adopt the assumption of favorable circumstances in whole. Stability – even stability for the right reasons – can be undermined by famine, war, disaster, etc. But to worry about these would be a distraction, forcing us to attend to cases of accidental instability, rather than the matter of inherent stability. For this reason, I leave the possibility of unfavorable circumstances out of this theoretical project.

The suitability of a strict compliance assumption for our purposes is more complex. On one hand, the precise formulation of Rawls’s assumption is inapplicable to the stability approach. Insofar as Rawls’s assumption concerns compliance with specifically just institution, and with laws that embody the two principles of justice as fairness, then it has no place in the approach I will pursue. There simply is no prior conception of “justice” in my account; we are interested in compliance with, or the stability of, whatever political order the regime happens to enshrine.

More importantly, compliance is of central importance to my account, and cannot be simply sidelined by means of an idealizing assumption. After all, subsequent chapters will not consider a regime’s capacity to motivate compliance as a mere side-constraint on the pursuit of some other aim (e.g. justice). Stability, for my purposes, is the aim. Instability is not just a strike against an otherwise desirable political principle. It’s the whole ball game. To assume full compliance, then, would be to assume away the basic problem that the theory is supposed to resolve. A meaningful investigation of SFRR’s requirements has to place the challenge of motivating compliance front and center.
On Rawls’s characterization, questions of citizens’ likelihood of complying with and supporting their regime is not precluded from consideration in ideal theory.\footnote{Simmons, “Ideal and Non-Ideal Theory,” 9–10.} Indeed, ideal theory ought to take into account relevant information about human psychology and social life under political institutions in order to avoid the construction of veritable ou-topias – “no places.” He suggests that the likelihood of compliance is a salient concern even in the original position, where parties must deliberate in cognizance of the “strains of commitment,” such that they will not select principles of justice that they will be unable to comply with.\footnote{Rawls, A Theory of Justice, 215.} Incorporating realistic constraints into ideal theory supports its basic purpose, which Rawls describes as to “probe[] the limits of practicable political possibility.”\footnote{John Rawls, Justice as Fairness: A Restatement, ed. Erin Kelly (Cambridge, MA: Belknap Press, 2001), 4.} Theorizing which aims at stability for the right reasons proceeds in this vein, taking into account facts of pluralism, social psychology, the formative influence of institutions, and so on.

All this said, I do adopt an idealizing assumption about compliance, though a weaker one than the assumption of strict compliance. Instead, the argument in subsequent chapters relies on an assumption of normal compliance. SFRR is not concerned with politically irrelevant acts of noncompliance, such as teenagers who experiment with illegal drugs, desperate persons who commit insurance fraud, or madmen who commit homicide. Consequential in practice as these may be, they do not reflect failures of the political order, but rather contingent factors (youth, poverty, psychopathy). They tell us nothing about the basic effectiveness of the regime in motivating compliance. Thus, I rely on an idealizing assumption to set aside these acts of noncompliance. The question is whether a regime can win support from healthy, non-desperate citizens acting in possession of and on their rational faculties.
In sum, while ideal theory does rely on some idealizing assumptions to help isolate the theoretical questions at hand, ideal theory need not be entirely insensitive to facts about the world.\textsuperscript{120} That the desideratum of stability requires attention to these constraints does not mean that it must fall into the camp of non-ideal theory.\textsuperscript{121} Stability for the right reasons is, as Rawls intended, a legitimate object of inquiry in ideal theory. The SFRR-based theory I will construct in subsequent chapters adopts one of Rawls’s assumptions in whole and another in part. The basic problem is one that cannot be assumed away, even within the bounds of ideal theory: how to motivate compliance in the face of profound and persistent reasons for citizens to noncomply? This is the question that will occupy the chapters that follow.

\textbf{2.9 Approximating Justice}

I have advanced several distinct yet related claims in this chapter. The first is that the stability secured by a modus vivendi is inferior – in the sense that it is less robust – to the kind of stability described by stability for the right reasons. The second is that Rawls’s notion of SFRR can be logically disentangled from his focus on justice and reformulated to describe citizens’ convergence around support for a regime. My third claim is that SFRR represents a reasonably realistic and fully desirable normative conception of stability in contexts of pluralism. Fourth, I argued that concern for stability is a reasonable component of normative theorizing insofar as neglecting stability means risking the distinctly bad consequences of disorder. Fifth, I situated SFRR in relation to the ideal/non-ideal distinction in political theory.

These claims amount to a defense of SFRR as a normative ideal – primarily against charges from a realist vantage of liberal “moralism.” If they are correct, they suggest that the

\textsuperscript{120} Cf. Cohen, “Facts and Principles.”
\textsuperscript{121} Cf. Robeyns, “Ideal Theory in Theory and Practice,” 349, where she claims that stability enters consideration only as an end to be traded off against others.
realist objections are directed to some degree at a caricature of liberalism that focuses myopically on rights-protection and distributive justice to the neglect of stability. This chapter has pushed back against the moralistic understanding of liberalism by focusing on the most influential recent formulation of the liberal position, challenging the notion that Rawlsian liberalism is unconcerned with stability. Moreover, it has positioned SFRR – a conceptual product of Rawlsian liberalism – as a viable normative response to the problem of stability.

Recovering stability as a central liberal concern is required by exegetical fidelity, but the tack I’ve taken might put some liberals at dis-ease. Does pursuit of stability independent of any substantive conception of justice mean neglecting the other commitments that liberalism holds dear, such as freedom, equality, and protection of rights? The arguments I have advanced depend on the moral neutrality of SFRR, but if this leaves us with a normative ideal that will sanction noxious political arrangements, we may be worse off than before. After all, the stability of a regime does not imply its justness, and it is eminently possible that stability might require transgressions against freedom, fairness, and so on. The political thinker who most definitively put stability first – Thomas Hobbes – did not defend a regime-type that would satisfy many liberals today. One might wonder: is following the realists in “putting stability first” bound to make liberalism less liberal?

First, on the assumption that stability and justice are non-overlapping, even potentially contradictory ends, it does not follow from what I have argued that stability must necessarily be the sole, or even the primary, normative standard for political theory. There is no reason that SFRR cannot be understood as one necessary condition for desirability among others, such as justice. Rawls’s theory demonstrates nicely how stability and justice might be pursued in

---

122 I cannot attempt it here, but a broader-scope historical project might profitably trace the role of stability in other key liberal texts. Along these lines, see, for example, Brandon P. Turner, “John Stuart Mill and the Antagonistic Foundation of Liberal Politics,” *The Review of Politics* 72 (2010): 25–53.
tandem, searching for principles of justice that are not only normatively satisfactory but also capable of achieving stability.\textsuperscript{123}

Second, while a conception of stability is not as overtly normative as a conception of justice, such a conception will nevertheless impose certain parameters defining the set of regimes that can achieve that version of stability. Thus, from a notion of stability is possible to derive political principles, at least in negative fashion – ruling out certain kinds of regimes as inherently unstable. Defenders of modus vivendi stability have done this, to varying degrees. For instance, Chandran Kukathas has argued that under a modus vivendi, the function of the state is to “preserve the order in which [diverse groups] can coexist … Its purpose is simply to preserve order so that people might live freely together.”\textsuperscript{124} The modus vivendi political order is a minimal one, curing the basic conditions of stable order. David McCabe advocates a slightly more robust role for the state, arguing that a modus vivendi is best served by a “minimal moral universalism,” such as a core set of rights which “draw the limits of the tolerable.” This floor of human well-being would rule out such pernicious practices as slavery, and would guarantee access to things like education, basic physical and psychological needs, and security.\textsuperscript{125} John Gray rejects the notion of human rights as immutable truths, but embraces a similar idea of basic minimal legitimacy, and above all emphasizes the norm of toleration. The goal of a modus vivendi is to secure the toleration of diverse ways of life in pursuit of diverse values.\textsuperscript{126}

\textsuperscript{123} Many have taken the view that the ability to be stably realized is a necessary condition of a satisfactory normative account of justice. Martha Nussbaum, for instance, writes that “Part of justifying a normative political project is showing that it can be reasonably stable.” \textit{Political Emotions: Why Love Matters for Justice} (Cambridge, MA: Belknap Press of Harvard University Press, 2013), 16. Also see Gilabert and Lawford-Smith, “Political Feasibility,” 813. For my purposes here, I treat the desideratum of stability as separate from what I am calling normative desirability.

\textsuperscript{124} Kukathas, \textit{The Liberal Archipelago}, 213.

\textsuperscript{125} McCabe, \textit{Modus Vivendi Liberalism}, 138.

\textsuperscript{126} Gray, \textit{Two Faces of Liberalism}. 
The ideal of SFRR is similarly capable of generating specific political principles that conduce to inherent, self-generating stability. It will be the project of the next two chapters to identify some of those principles. For the moment, a general observation will suffice. In order for a regime to win the endorsement of most or all of its subjects, it must serve those persons’ interests reasonably well, and must treat them with some significant degree of procedural fairness. The basic idea is intuitive: when people feel that their government treats them badly or arbitrarily, they are less likely (absent some coercive reason) to support it.\textsuperscript{127} A regime that can achieve SFRR will not be one that imposes egregious harms on its citizens or discriminates radically among them.

It is not surprising that the political principles derived from SFRR should be broadly compatible with common conceptions of justice or liberal sensibilities in general, given that SFRR rests on the justification of a regime to the citizens who live under it. In this respect, SFRR crystallizes the conception of stability underwriting the social contract tradition as a whole. The requirement of justification will mean that any nasty regimes that treat citizens poorly will be excluded: they are incapable of becoming stable for the right reasons. Gaus notes this fact in addressing the concern that a convergence conception of justification will sanction illiberal political arrangements:

\begin{quotation}
“The constitutions that really will be vetoed are those that build social order on the requirement that some renounce their most cherished convictions even as the ideals of their own personal lives. While many of us many rank most highly social orders that
\end{quotation}

\textsuperscript{127}Hobbes provides the most obvious guide here. He held that in order to maintain a stable commonwealth, the sovereign must not only preserve the lives of his subjects, but also treat them in accordance with “equity” and the rule of law. \textit{Leviathan}, chap. 15, pp. 96–97.
express our ideals in public life, what is truly unacceptable are social and political orders that insist we abandon our deep ideals as the basis of our own existence.”

A regime that can become stable for the right reasons will not be one that imposes egregious harms on its citizens or discriminates radically among them. On the other hand, some unpleasant arrangements might be consistent with SFRR – such as oppression of very small minorities or harms supported by internalized justifying ideologies. These possibilities are, of course, unsatisfactory from a liberal point of view. My contention is not that SFRR is an adequate substitute for a normative standard of justice, but rather that SFRR may achieve an approximation of justice by ruling out at least the worst kinds of offensive regimes.

Though I have pursued this final line of argument only suggestively, the possibility that SFRR might describe a regime that is both reasonably likely to endure and substantively conducive to desirable forms of government would constitute another mark in its favor as a normative ideal. Normative desirability and longevity are each necessary but independently insufficient for a regime to qualify as choiceworthy. If what I have argued here is correct, then stability for the right reasons actually represents the alignment of these twin aims in a single goal that entails at least the partial satisfaction of both. For this reason, among the others advanced here, it is the conception of stability that ought to be preferred – by liberals and realists alike.

---

CHAPTER THREE:

FROM STABILITY TO IMPARTIALITY: A HOBBESIAN ACCOUNT

In the last chapter, in explicating and defending the normative ideal of stability for the right reasons, we considered the relationship of that ideal to the fact of pluralism. I argued that stability for the right reasons neither assumes nor prescribes homogeneity, and is indeed compatible with a robust pluralism of comprehensive doctrines. That free societies are, and will inevitably, be characterized by a diversity of beliefs, at least some of which are incompatible with one another, need not rule out inherent, non-coerced stability as a realistic and desirable goal for political societies. This is not to say, however, that achieving stability for the right reasons will not be difficult, nor that this difficulty is not amplified by the fact of pluralism. Stability for the right reasons sets a high bar, and not all regimes will meet the standard. That citizens disagree about many important things means that identifying those political principles around which they can converge in their support will likely be a challenge – or at least more challenging than if citizens all agreed on most important matters.

It is the objective of this chapter to identify some of the necessary conditions for the possibility of stability amid conditions of pluralism. (Unless otherwise noted, references to stability should be understood to mean stability for the right reasons.) More specifically, my aim is to show that the liberal norms of impartiality – distributive and justificatory – area among those necessary conditions. These are required in order to supply citizens with sufficient reasons to maintain their commitments to support the regime, and thereby to overcome pervasive challenges to stability.

The chapter proceeds in several steps. Section 3.1 accomplishes some stage-setting by exploring what it means to take stability as an end from which to derive political principles. This
is done by way of contrast with the “liberalism of fear” advanced by Judith Shklar and “negative morality” more generally. Section 3.2 further prepares the ground by considering how stability for the right reasons compares to the conception of stability at work in Hobbes. In section 3.3, we examine the substantive challenges that pluralism poses to the stability of a political regime, relying on Hobbes’s insights on the conflict-generating potential of disagreement and competing group loyalty. Next, in 3.4, we identify four instances of impartiality in Hobbes’s theory of the civil commonwealth. These each provide some presumptive support for the notion that political impartiality is at least consistent with stability, and suggest some resources for an original defense of impartiality on the basis of stability. Sections 3.5 and 3.6 contain the primary constructive material of this chapter, supplying arguments from stability to distributive and justificatory impartiality, respectively. These arguments show, without appeal to commitments such as equality or freedom, that impartiality is a necessary (if not sufficient) condition for the stability of a political order. The conclusion of these arguments prepares the way for chapter 4, where we will consider the motivational challenges faced by impartialist regimes.

The previous chapter focused on the conception of stability that John Rawls affirmed, and I have alluded in a couple places to the role stability played in his work. As scholars have come to appreciate, stability was a major concern for Rawls, and motivated the shift from the earlier to later formulations of his theory.¹ Rawls’s arguments from stability might have played a prominent role in this chapter, as they provide a way to ground impartiality in the end of stability. Rawls’s role here is more minimal, though. Instead, I look further back, to Thomas Hobbes, for the principal resources in developing my arguments. This is for the sake of isolating stability as the motivating concern. As mentioned in the last chapter, Rawls’s work is focused on

---

the twin aims of justice and stability, and it can be difficult to tease these two commitments apart. Hobbes, in contrast, eschews any independent standard of justice apart from that which conduces to civil peace. In order to make clear that the original arguments advanced here also set justice aside, we will take Hobbes as our primary guide, rather than Rawls.

3.1 Fear and Hope

Before proceeding to the constructive account of pluralism and impartiality I will advance, I want to situate my approach in relation to an important style of thinking about politics that been somewhat influential in recent decades. Whereas political theory is often – and has since antiquity – been articulated in terms of the positive ends that the political order might strive to achieve or secure, it is also possible to take the opposite approach, focusing on the things that politics might strive to avoid or to mitigate. The classic contemporary articulation of such an approach is found in the work of Judith Shklar, particularly her famous essay titled “The Liberalism of Fear.”2 On Shklar’s account, liberal politics takes as its basic and primary function the mitigation of fear. The fear she is concerned about arises from the threat of cruelty. Liberal politics must “put cruelty first,” meaning that it must strive first to eliminate the threat of cruelty in the lives of its subjects. This has been described as a “negative” approach to politics, in the sense that it focuses not on positive ends to be sought, but rather on the summum malum which is to be avoided.3 In the words of Michael Walzer, “‘Putting cruelty first’ means starting with what we most want to escape.”4

---

It might seem that the approach I am adopting here, prioritizing stability rather than other substantive moral commitments, would have some significant affinity with political theory in the “negative” mode. The desirability of stability stems in large part from the undesirability of instability and disorder. The *locus classicus* for painting stability as a flight from great evil is, of course, Hobbes’s *Leviathan*, which grounds political order entirely on its ability to rescue persons from the violent conflict of the state of nature. Stability, on this Hobbesian view, is defined precisely by “what we most want to escape”: disorder and the vulnerability it entails. The apparent affinity between a stability-based approach to politics and the negative politics of fear is further supported by the fact that both approaches eschew “thick” positive ends, focusing rather on mitigating certain evils and thereby creating space for persons to pursue the ends that they choose. This is why Shklar’s approach is a variety of *liberalism*: it ensures citizens’ freedom from cruelty and fear. This is a decidedly negative (in the Berlinian sense) form of liberty, insofar as it does not advance any positive conception of human good.\(^5\) Similarly, to prioritize stability is to set aside questions of the good life or human flourishing, or at least to relegate them to secondary importance. Stability is not a perfectionist end: it is unconcerned with how citizens choose to live, so long as those ways of life are consistent with stability.

Jacob Levy has applied the lens of fear to the challenge of pluralism in his 2000 book *The Multiculturalism of Fear*. Levy extends Shklar’s argument to include humiliation as a relative of cruelty, concluding that the logic of the liberalism of fear requires attending to the way that persons of diverse identities, beliefs, and group membership relate to one another in a pluralist society. As is all too well-known, these kinds of loyalties kindle the possibility of conflict,

\(^5\) Shklar, “The Liberalism of Fear,” 28–31. Shklar distinguishes her view from Berlin’s by noting that the liberalism of fear does require attention to the “conditions of liberty” – the social and political institutions that make freedom possible. But the two views overlap in their anti-perfectionism. As Shklar writes, “No form of liberalism has any business telling the citizenry to pursue happiness or even to define that wholly elusive condition” (31).
humiliation, and – at the most extreme – violence. Levy writes, “The social facts about nationalism and multiculturalism … generate social situations in which the fears of a liberalism of fear may be realized. They provide the opportunity for political violence and cruelty.”

Identities, beliefs, and memberships are of significance for politics “because of what they risk doing to common social and political life.” Political theory, Levy argues, must be attentive to individual identity and group membership for the reason that these features can be potent generators of cruelty and fear. In a similar vein, my aim in this project is to consider the ways that pluralism can support or challenge political stability. The *summum malum* in this case is not cruelty, but rather disorder, and insofar as the presence of diverse beliefs and groups can undermine stability, fear of that basic evil urges attention to pluralism. In this respect, my project bears a distinct similarity to Levy’s and, although the connection is less immediate, Shklar’s.

That said, there are a couple important differences between my project and the theories of fear advanced by Shklar and Levy. First, to adopt the aim of stability for the right reasons is to do more than hope for the mitigation of disorder. Of course, to desire P often corresponds to a desire to avoid not-P, and vice versa, so the distinction between desiring stability and wishing to avoid instability might be an elusive one. But stability for the right reasons describes a state defined by more than the mere absence of disorder. It is characterized by order of a particular kind, which we described in the previous chapter. Stability for the right reasons constitutes a positive conception of stability at which to aim. To the extent that the end of stability does entail a desire to avoid instability, this desire does not rest on the identification of any specific *summum malum*. The reasons for preferring stability to instability are multiple, and this account takes no stand on which among them (if any) is worst or most aversion-worthy. The account beginning

---

7 Ibid., 33.
from stability is not strictly a “negative” approach to politics, as it neither identifies the thing-to-be-escaped nor focuses solely on the mitigation of that thing.

Additionally, my approach differs from the liberalism of fear insofar as the principles to be derived from the end of stability do not necessarily constitute a ‘liberalism.’ That is, there is no guarantee (nor even any obvious a priori reason to believe) that stability will require political rules and institutions that promote liberty or protect rights. It is my ultimate aim to show that the norms that stability requires do indeed track some prominent liberal commitments closely, but this is a conclusion to be established, not assumed or asserted at the outset. In contrast, Shklar and Levy are clear that their accounts represent interpretations of liberalism: avoidance of cruelty is a way to account for the core protections that liberal theories afford. Insofar as the end of fear-mitigation fails to ground some important liberal commitments, it must be supplemented by a more robust, positive conception of liberty. Amy Gutmann makes this point in arguing that a purely negative liberalism cannot ground the core liberal principle of toleration, which requires positive concern for personal integrity and individual conscience, not merely negative protection from harms.8 Similarly, Levy admits that the negative goal of fear-mitigation cannot ground all of the commitments that make up a satisfactory political theory. Rather, the multiculturalism of fear must be accompanied by the multiculturalism of rights.9 I have no objection to a political theory grounded in multiple distinct ends, but my task here is to see how far stability alone can carry us.

My project is therefore at once both more and less front-loaded than the liberalism of fear. In beginning from the aim of SFRR, my approach adopts a more robust, positive conception

9 Levy, The Multiculturalism of Fear, 37.
of stability than the strictly negative orientation of Shklar and Levy. But whereas their accounts do aim at rights-protection and some guarantee of liberty, mine eschews commitments like these. Though I hope to show that some liberal commitments can be derived, the only normative “input” in my account is the desirability of a certain conception of stability.

In some respects, the stability approach is more accurately characterized as hopeful, rather than fearful. It is hopeful in that it reflects confidence that there exist political principles under which persons could live together in stable conditions grounded otherwise than in coercion – and moreover that it is possible to identify these principles. The alternative is to assume that only by force can an enduring concord be achieved among humans. While many would accept this claim without hesitation, one cannot deny the pessimism it entails. To begin from the goal of stability for the right reasons and to seek the political principles necessary to support it is to hope in a different possibility: a peace enforced not by the sword but by the will of those who constitute it. Alexander Hamilton posed the salient question in the opening question of Federalist No. 1: “whether societies … are really capable or not of establishing good government from reflection and choice, or whether they are forever destined to depend for their political constitutions on accident and force.”\textsuperscript{10} I alluded in the last chapter to the significance of this hope for Rawls, who, at the end of \textit{The Law of Peoples}, described the basic import of his theorizing about justice in stark terms: “If a reasonably just [society] is not possible … one might ask, with Kant, whether it is worthwhile for human beings to live on the earth.”\textsuperscript{11} I take no position on the worthiness of human life, but I share the conviction of both Publius and Rawls that a non-

\textsuperscript{10} \textit{The Federalist} #1, quoted in Fred D’Agostino, \textit{Free Public Reason: Making It Up As We Go} (New York: Oxford University Press, 1996), 10.

coerced peace is preferable to one secured by coercion. It is with the hope of identifying the conditions for such a peace that we proceed.

3.2 Hobbesian Stability

I have said that the account I tend to advance in this chapter is a Hobbesian one, and some comments are necessary to clarify the relationship of Thomas Hobbes to the argument that is to follow. First, in offering a Hobbesian argument, I do not mean to suggest that my primary orientation is exegetical. I am not concerned first and foremost to identify Hobbes’s views or to determine what Hobbes intended in this or that text. Rather, I mean to indicate that in Hobbes’s work there are resources that can be helpfully appropriated for the purpose of addressing the issues at hand in this project. Some of Hobbes’s insights, claims, and logical maneuvers provide a useful framework on which to build my argument. But I rely on Hobbes in this way with no desire to attribute any aspects of my argument to him. My argument is Hobbesian in the sense that it is inspired by Hobbes, not necessarily that it is one that Hobbes did or would himself affirm.

Hobbes is an obvious figure to invoke in consideration of stability, and he has already made appearances in this dissertation so far. More than any other canonical thinker, Hobbes takes the problem of stability seriously. The necessity of establishing a stable political order and the challenges that confront such an endeavor sit at the center of his political thought. At the same time, Hobbes’s role in the previous chapter was to serve as a foil, providing a contrast to the conception of stability for the right reasons. Whereas SFRR insists on non-coerced stability, Hobbesian stability is emphatically coercive: it is premised on the force of the sovereign. “Covenants without the sword are but words, and of no strength to secure a man at all,” Hobbes
writes.\textsuperscript{12} What S. A. Lloyd has called the “standard philosophical interpretation”\textsuperscript{13} of Hobbes’s theory takes the coercive function of the sovereign as basic. Gauthier, for instance, writes that “the basic right of sovereign is the right to exercise the two swords, of justice and war,” the first of which motivates obedience to the sovereign, the second of which defends the commonwealth against external threats.\textsuperscript{14} The centrality of coercion in this interpretation of Hobbes would suggest that Hobbes’s conception of stability is fundamentally unlike the one I am employing here. Is he not, then, an inappropriate resource for my argument?

While Hobbes is unquestionably more comfortable with the use of coercion as a mechanism to achieve stability than is the ideal of stability for the right reasons, it is not clear that even Hobbes endorsed an ideal of stability grounded \textit{purely} in coercion. For one thing, Hobbes simply did not think it \textit{possible} to ground stability in coercion alone. In chapter 30 of \textit{Leviathan}, Hobbes asserts that “the grounds of [the sovereign’s] rights have the rather need to be diligently, and truly taught; because they cannot be maintained by any civil law, or terror of legal punishment.”\textsuperscript{15} The problem for Hobbes is that, as Lloyd notes, “People simply are not the bodily preservation-centered egoists needed to make [the threat of punishment] motivationally reliable. … They have interests that may \textit{transcend} their interest in temporal bodily preservation, interests for the sake of which they may be willing to risk death, or even to embrace death.”\textsuperscript{16} In contrast to the claim of the standard philosophical interpretation, Hobbes recognized that order could neither be achieved nor maintained on the basis of coercive force alone.

\begin{footnotesize}
\begin{enumerate}
\item David Gauthier, \textit{Logic of Leviathan} (New York: Oxford University Press, 1979), 108.
\item Hobbes, \textit{Leviathan}, chap. 30, par. 4, p. 220, emphasis added.
\end{enumerate}
\end{footnotesize}
An alternative reading of Hobbes holds that even if his commonwealth is not maintained purely by coercive force, it nevertheless represents a more strategic form of stability than that described by stability for the right reasons. John Gray, for instance, describes the Hobbesian commonwealth as a quintessential modus vivendi, writing that “Hobbes is thereby the progenitor of a tradition of liberal thought in which modus vivendi is central.” Gray affirms that a modus vivendi is the appropriate response to the fact of pluralism. Rather than hope for any meaningful consensus on political values or principles, a modus vivendi keeps the universal claims of its subjects at arm’s length, and seeks to establish merely a compromise framework of rules and institutions within which citizens can live peacefully. This is, according to Gray, what Hobbes envisioned. “A Hobbesian state,” he writes, “extends to private belief the radical tolerance of indifference.” The commonwealth is not the embodiment of any particular substantive value system, but rather a mechanism that all can accept for the management of competing beliefs and commitments. This interpretation of Hobbes’s aims explains his comfort with coercion: the goal is to secure a peaceful compromise by whatever efficacious means are available.

This interpretation of Hobbes is inaccurate. A modus vivendi is a political arrangement determined by compromise that lacks any justificatory basis in the private reasoning of citizens. Gray’s claim is that the Hobbesian commonwealth satisfies this definition by treating citizens’ private beliefs with indifference. To the extent that the commonwealth is understood as simply a prudential arrangement which persons abide for the sake of stability, this characterization may seem correct. But Hobbes is actually quite concerned with the content of persons’ private beliefs; he regards them with anything but indifference. As we will explore below, Hobbes both identifies divergent private beliefs as a principal cause of conflict, and suggests that the

---

18 Ibid.
commonwealth must strive to reconcile those beliefs to the maintenance of public order. Hobbes writes that “the actions of men proceed from their opinions, and in the well-governing of opinions consisteth the well-governing of men’s actions, in order to their peace and concord.”

The Hobbesian commonwealth cannot be understood as a neutral framework of laws and institutions. It is more than a modus vivendi. It is a mechanism for achieving a workable harmony (if not outright unity) among the private beliefs of its citizens.

That Hobbes believes political stability to require intervention in the beliefs of citizens is evinced by the prominent role he grants to education in the civil commonwealth. Hobbes judged it necessary for stability that a regime inculcate beliefs and attitudes supportive of the regime. Among the strategies that Hobbes invokes in order to maintain compliant dispositions among subjects is to permit the teaching of doctrines that might undermine the authority of the sovereign, but not “without present applying such correctives of discreet masters, as are fit to take away their venom.” A further function of education, on Hobbes’s account, is in part to discourage subjects from blaming the sovereign for their woes (and consequently developing a desire to resist the sovereign), and also to persuade subjects that their continued compliance with the sovereign’s requirements is, on the whole, a good thing for them. That is, education serves to make Hobbesian subjects aware of the reasons they have to endorse the regime under which they live.

19 Hobbes, Leviathan, chap. 18, par. 9, p. 113.
That Hobbes affirms the value of shaping citizens’ perceptions in this way suggests that Hobbesian stability aims to be self-supporting without relying exclusively on coercive force. S.A. Lloyd captures this feature of Hobbes’s project in terms that suggest its similarity to the notion of inherent stability:

Hobbes is aiming to identify a principle of political obligation that could, if followed, ensure the perpetual maintenance of effective social order (that is, ensure a commonwealth that could never be destroyed but by foreign war), and that nearly everyone, given what each acknowledges to be his interests, can have sufficient reason to affirm and uphold.\(^\text{24}\)

I am suggesting that the Hobbesian project is not so fundamentally opposed to the kind of political aspiration described by stability for the right reasons. A number of authors have gone further, emphasizing even more strongly and directly the Hobbesian character of Rawls’s theorizing in general. Duncan Ivison has identified a “secret history” of public reason that leads to Rawls but begins not with Kant (as is typically assumed) but rather with Hobbes.\(^\text{25}\) John Gray, in his 1993 review of Rawls’s Political Liberalism goes so far as to assert that Rawls’s ideas “are not new. They are stated with unsurpassed clarity by Thomas Hobbes, at the start of the modern period, in his masterpiece, Leviathan.”\(^\text{26}\) Even Rawls himself apparently noted the affinity, as S. A. Lloyd recounts: “Rawls has told me that he suspects … that Hobbes may have been the first

\(^{24}\) Lloyd, *Ideals as Interests in Hobbes’ Leviathan*, 51.
political liberal, and he has asked me to investigate this suggestion.”27 Could it be, then, that Rawls has been widely misunderstood as a neo-Kantian, when in fact his justice as fairness represents a (somewhat veiled) revival of Hobbesianism?28

This is too strong a claim. Even though there are significant areas of affinity between Hobbes’s project and Rawls’s, there remain points of fundamental contrast which preclude any attempt to join the Leviathan and justice as fairness too closely. Lloyd has pushed back against the notion of a Hobbes-Rawls alliance, arguing that Hobbes’s insistence on bringing comprehensive doctrines into harmony with the reigning social order is incompatible with the strategy of Rawls’s political liberalism, which involves transcending pluralism by grounding political principles in an act of abstraction (the veil of ignorance).29 For my purposes, I do not need to assume anything like the strong claim that Rawls is basically a Hobbesian (or vice versa). I draw the comparison only in order to defend my invocation of Hobbes as a resource for working out the requirements of the Rawlsian conception of stability for the right reasons. There is sufficient overlap between Hobbes’s objectives and the goal of stability for the right reasons (even if this overlap is not exhaustive), that we may profitably draw resources from the former to assist in our analysis of the latter. This is the extent of the interpretive claim I mean to stake on the relationship between Hobbes and Rawls.

3.3 Stability and the “Perils of Pluralism”\textsuperscript{30}

According to the “standard philosophical interpretation,” Hobbes’s diagnosis of the problem of order can be modeled in game theoretic terms as a problem arising from the rational actions of self-interested individuals. Persons in the state of nature, assuming rationality and an overriding desire for self-preservation, find that even if cooperation were to their ultimate benefit, acting individually to initiate cooperation would be suicidal. On this view, the state of nature becomes a state of war because of the structural features of anarchy and the innate features of human nature. Modeling humanity in this way implies that individuals are fundamentally alike and are situated symmetrically with regard to one another, and thus the particular qualities of one or another are not ultimately decisive.\textsuperscript{31} If this is true, then understanding the dynamics of humanity’s pre-political condition requires no consideration of particular persons’ individual characteristics, commitments, or beliefs.

There are some merits to this reading of Hobbes, but it grossly overlooks the psychological and sociological nuances of Hobbes’s theory. For instance, the influence of “vain glory” on persons’ desires and motivations features prominently throughout the \textit{Leviathan}. Or, as mentioned just above, his attention to education suggests that individuals’ \textit{beliefs} are of significance for Hobbes’s understanding of conflict. Similarly with respect to religion, which occupies a great deal of Hobbes’s attention. In contrast to the rational choice interpretation, Hobbes’s theory is sensitive to the ways in which qualities of particular persons – and the ways in which these qualities differentiate persons from one another – can create conflict and


\textsuperscript{31} For instance, Gauthier writes that “The prudent engine will always seek to secure itself, to the maximum extent possible, and so will desire ‘power after power’ as long as vital motion continues. … For if your power may be opposed to mine, may hinder my efforts to maintain my vital motion, then I must strive to secure myself by increasing my power in relation to yours. You, however, must equally strive to secure yourself by increasing your power in relation to mine. And so the race is on.” \textit{Logic of Leviathan}, 14.
challenge the achievement of order. On the purely structural model, the fact of pluralism would be insignificant: whether people agree or disagree, they are prone to conflict. More nuanced and attentive readings of Hobbes, however, reveal him to be eminently concerned with the destabilizing potential of pluralism. His concerns in this regard can helpfully illuminate the challenge of pluralism that must be overcome if stability is to be achieved.

The first step in appreciating Hobbes’s treatment of pluralism involves recognizing that the establishment of the commonwealth does not, on Hobbes’s view, resolve the problem of disorder. One way to read the *Leviathan* is to focus on the obstacles to creating order in the state of nature. Some interpreters have argued that the state of nature is properly understood as a generalized prisoner’s dilemma, where each person’s welfare would be maximized by mutual cooperation, but given uncertainty about others’ intentions, defection remains each’s dominant option. The sovereign provides a solution to the dilemma by altering individuals’ payoff tables by means of coercive sanctions: defection, once subject to punishment by the sovereign, carries a penalty sufficient to render this option unattractive. With persons’ balances of reasons altered thusly, cooperation becomes both individually and collectively rational, and civil society is formed.

This cannot be the whole story, however. If the sovereign’s function were merely to resolve the prisoner’s dilemma inherent in the state of nature, the establishment of sovereign authority would dispatch the problem of disorder once and for all. The maintenance of coercive

---


33 Gauthier, *Logic of Leviathan*, 86.
penalties would ensure that no citizen had an incentive to defect, and cooperation among subjects would endure. An attentive reading of Hobbes makes clear, though, that he worried about threats to the maintenance of order within the commonwealth. Writing against the backdrop of the English Civil War, Hobbes’s aim was not merely to describe the path from anarchy to order, but to identify the conditions necessary to avoid the dissolution of that order. There is no reason, Hobbes thinks, that a civil order cannot endure “as long as mankind, or as the laws of nature, or as justice itself.” When a regime collapses, the fault lies in its inadequate design.\textsuperscript{34} The real challenge, then, is to determine the principles of political order that are sufficient to overcome the persistent challenges to stability. As Arash Abizadeh notes, Hobbes’s real question “is not how to enter political society – this could be done under the right circumstances when fear of death prevails – but how to stay in political society.”\textsuperscript{35}

There are, then, three distinct categories of problems that one can identify for Hobbes.\textsuperscript{36} The first are the causes of war: those features of human psychology, material reality, or the structure of anarchy that make the state of nature inevitably a state of war. The second are obstacles to peace: those facts that impede the transition from anarchy to civil society. The third are threats to stability: those features of individual and social psychology that can potentially undermine the authority of the sovereign – and, consequently, civil order itself. Our focus here is on the third class of problems. Apart from how an order comes about, what are the facts which threaten to undermine it? It is possible, though not necessary, that a factor might play all three of these roles. For instance, two of the explanations Hobbes invokes to explain why the state of nature is a state of war – competition over scarce resources and “diffidence,” or fear of others’

\textsuperscript{36} I draw here, with modifications, on Abizadeh, “Hobbes on the Causes of War,” 301–2.
intentions – cannot explain the vulnerability of established order. They are problems that should be resolved by the coordinating function of a sovereign authority, and thus will not pose an ongoing threat to the endurance of that authority. In contrast, the third factor Hobbes identifies as contributing to the state of war – glory – may indeed be an ongoing problem, even within civil society. (More on this below.)

Whatever the full set of facts that might pose a threat to the maintenance of civil society, it is clear that for Hobbes among the most worrisome are those arising from differences among persons. For Hobbes, there are at least two ways that pluralism threatens to generate conflict. The first concerns the simple fact of disagreement, while the second concerns the tendency of persons to group themselves into voluntary associations.

Disagreement

On Hobbes’s view, it is not only competing interests that generate conflict among persons, but also competing beliefs. He notes that the tempers, customs, and doctrines of men are different; and divers men differ not only in their judgment on the senses (of what is pleasant and unpleasant to the taste, smell, hearing, touch, and sight), but also of what is conformable or disagreeable to reason in the actions of common life... from whence arise disputes, controversies, and at last war.37

In short, differences in judgment give rise to disagreement and, ultimately, war. It might be thought, though, that this is a contingent feature of human interaction. Perhaps differences in judgment are the result of faulty exercises of reason, and could be prevented by correcting these

37 Hobbes, Leviathan, chap. 15, par. 40, p. 100.
failures. So corrected, individuals would reason to identical conclusions. If this is the case, then prevention of conflict (for reasons of disagreement) is a matter of properly informing the reasoning faculties of individuals. Hobbes recognizes that the fallibility of reason can produce disagreements, but this is not their only source. Rather, Hobbes recognizes that disagreement can arise from differences in appetites and the judgments that concern them. It is obvious that persons differ in their preferences and desires, but Hobbes recognizes additionally that even persons of identical appetites will judge differently how best to satisfy those appetites in some given actual situation. Hobbes writes,

Whatsoever is the object of any mans appetite or desire that is it which he for his part calleth good; and the object of his hate and aversion, evil; and of his contempt, vile and inconsiderable. For these words of good, evil, and contemptible are ever used with relation to the person that useth them, there being nothing simply and absolutely so, nor any common rule of good and evil to be taken from the nature of the objects themselves, but from the Person of the man…

Gauthier explains Hobbes’s view in the following example: “You and I may both find praise pleasant and criticism unpleasant; we shall therefore evaluate quite differently a state of affairs in which you are praised and I am criticized. You may seek to bring about this state of affairs, whereas I seek to prevent it.” Persons are bound to differ in how they judge the desirability of a state of affairs, even when their appetites themselves are identical. This fact, along with the fallibility of reason and the divergence of appetites, underwrites what Hobbes sees

39 Hobbes, Leviathan, chap. 6, par. 7, p. 28–29.
as the inevitability of disagreement. In this view, Hobbes agrees with Rawls, who cites the
“burdens of judgment” as sources of unavoidable disagreement. The fact of life in a free society
is that “it is not to be expected that conscientious persons with full powers of reason, even after
free discussion, will all arrive at the same conclusion.”\footnote{John Rawls, \textit{Political Liberalism} (New York: Columbia University Press, 1996), 58.} The challenge for political society, then,
is not mere disagreement, but what Rawls called “reasonable disagreement,” which arises not
from error or faulty reason, but is an inevitable outcome even when reason is exercised properly.

How, though, does disagreement lead to conflict? The move from differences in
judgment to “disputes, controversies, and at last war” cannot be simply assumed. There must be
some mechanism or process by which disagreement creates discord. Interpreters have attributed
to Hobbes a couple possible accounts.

One version, advanced by Richard Tuck and Sheldon Wolin, holds that Hobbes saw
conflict arising from disagreement because disagreement entails the absence of a common
normative vocabulary that can ground shard understandings of the norms that ought to guide
existence unless they were capable of using a common moral language.”\footnote{Tuck, \textit{Hobbes}, 55.} Though this belief is
assumed in Hobbes, it is not clearly explained or defended. Tuck goes nearly so far as to claim
that vocabulary conflicts in the state of nature are actually reflective of the normative conflict
between individuals’ opposing rights, though this normative reading has the effect of sidelining
Hobbes’s psychological account of the causes of the state of war.\footnote{Ibid., 107–8; see Abizadeh, “Hobbes on the Causes of War,” 307.}

An alternative connection between the lack of shared normative vocabulary and conflict
focuses on the role that language plays in adjudicating conflicts. Without access to shared
understandings, it will be impossible to resolve the clashes that inevitably arise among persons (including from sources such as competition and diffidence). Sheldon Wolin writes that Hobbes operated on the recognition that “a political order involved more than power, authority, law, and institutions: it was a sensitive system of communication dependent upon a system of verbal signs, actions, and gestures bearing generally accepted meaning.” Without shared understandings of terms such as “liberty,” “authority,” or “property,” persons are unable to effectively coordinate their actions, and are bound to find themselves in conflict. In order for a system of rules to function harmoniously, persons must engage it on common terms. On this view, then, the sovereign’s role is that of the “Great Definer” who establishes an objective set of shared terms to ground interpretation of the law and resolution of conflicts.

Arash Abizadeh has offered an additional account of how disagreement leads to conflict in Hobbes. His view draws on one of Hobbes’s central psychological considerations: glory. The influence of vainglory in prompting conflict between persons each desiring admiration and honor (over against the other, of course) has been explored most prominently by Leo Strauss. Abizadeh focuses more specifically on how a drive for glory can create conflict from ideological disagreement. According to his view, humans are sensitive to any slight or perceived expression of contempt from others. Because of the strong way that persons identify with their beliefs, they tend to perceive others’ disagreement as expressions of contempt. “Prickly humans feel contempt by another’s different opinion because of their natural tendency to identify with their own opinions, that is, to see their opinions as constitutive of who they are—no matter how trivial

45 Wolin, Politics and Vision, 231.
46 Ibid., 232.
or unfounded those opinions happen to be. They take disagreement personally.48 As Hobbes notes, “To agree with in opinion is to honour, … To dissent is dishonor and an upbraiding of error.”49 Persons perceiving such offenses are prone to respond strongly and combatively. Hobbes believes also that disagreements over normative language, including political terms as well as theological and liturgical concepts, are most likely to give rise to debates and conflict. In the absence of shared understandings, then, persons are likely to enter into conflict as a result of the slights to each’s glory entailed by their disagreements.

Hobbes thus suggests a couple different ways in which disagreement might lead to conflict among persons. In moments of severe conflict, civil order is disrupted and stability is frustrated. There is an additional way in which disagreement can undermine stability – one that has nothing to do with outright conflict. In the last chapter, we identified the requirement that, in order for a political principle to be stable it must provide each citizen (or some sizeable portion of citizens) with a sufficient reason to endorse it, and citizens must recognize those reasons. That is, if citizens do not, from their own perspective, recognize a reason to support a political arrangement, they may be inclined to defect. In conditions of ideological homogeneity, achieving stability in this respect would be relatively easy, as it would involve shaping political laws and institutions that are supported by some belief held by each member of the populace. In conditions of pluralism, however, the task becomes more difficult. As a society becomes more pluralistic, the set of beliefs held in common by all citizens will tend to shrink, as will the set of political principles they can all affirm. This is not to say that these sets must necessarily shrink to nothing as pluralism increases, but the more disparate individuals’ views grow, identifying these sets (if they exist) will become ever more challenging. Thus, even if disagreement does not lead to

49 Hobbes, Leviathan, ch.10, par. 30, p. 52.
outright conflict between citizens, it can increase the difficulty of identifying political
arrangements that can win the support of a sufficient number of citizens as to endure over time.
In this way, too, disagreement poses a challenge to stability.

*Groups*

Hobbes recognizes that the challenge posed by pluralism extends beyond individuals and
their disagreements. It also includes—and perhaps takes strongest form in—the relations of
competing associations within society. Just as relevant ideological differences can be easily
overlooked in interpreting Hobbes, so can the significance of group identity. After all, Hobbes is
(rightly) recognized as a determined individualist, taking the socially alienated individual as the
basic unit of society and of political analysis. Any association, including the state, is for Hobbes
an artificial arrangement, possessing no intrinsic being or dignity. Moreover, in the Hobbesian
commonwealth the state is the sole authoritative association, with all others existing only at the
grace of the sovereign. In light of these facts, it might seem that Hobbes is exclusively—
perhaps even myopically—focused on persons as individuals, apart from whatever associations
they may tend to form. But this is to overlook the sociologically sensitive nuances of Hobbes’s
view. Specifically, it is to neglect Hobbes’s distinct attention to—and wariness toward—private
associations.

---

51 Hobbes describes the commonwealth as an “artificial man, though of greater stature and strength than the natural,
for whose protection and defence it was intended” (Introduction, par. 1, p. 3). In *De Cive*, Hobbes writes that “We
do not therefore by nature seek society for its own sake, but that we may receive some honour or profit from it.”
(chap. 1, sec. 2).
52 “And the sovereign, in every commonwealth, is the absolute representative of all the subjects, and therefore no
other can be representative of any part of them, but so far forth as he shall give leave.” *Leviathan*, chap. 22, par. 5, p.
146.
Hobbes’s principal concern is that private associations pose a threat to sovereign authority. They do this first by gathering persons in service of some joint project which may override their individual interests. Persons acting as members of a group are prone to do things—irrational things—that an individual acting alone would never dare. Hobbes writes,

Though the effect of folly in them that are possessed of an opinion of being inspired be not visible always in one many by any very extravagant action that proceedeth from such passion, yet when many of them conspire together, the rage of the whole multitude is visible enough. … For they will clamour, fight against, and destroy those by whom all their lifetime before they have been protected and secured from injury.  

Hobbes was concerned that associations not only serve to inflame passions and inspire irrational action from which individuals would ordinarily refrain, but additionally that groups are prone to engage in conflict with those around them. This is so even when those neighbors have heretofore served as able and civil partners in the maintenance of order. Persons acting on their own are, at most, a minimal disruption, but many persons “conspiring together” present a different order of threat.

More worrisome than setting person against person, however, is the tendency of associations to establish objects of loyalty other than the sovereign. Citizens may direct their loyalties at the causes to which their associations are devoted, or to the other members of their collective, or to the leaders of their respective groups. Hobbes recognizes the proclivity of persons to be influenced by “the preachers, or the most potent of the gentlemen that dwell

54 Hobbes, *Leviathan*, chap. 8, par. 21, p. 42.
amongst them,”55 and suggests that the “popularity of a potent subject” is a “dangerous disease,” for the reason that it causes people to be “drawn away from their obedience to the laws.”56 Humans are inclined to attach themselves to collectives, and both the aims of such collectives and the individuals who constitute or direct them provide alternative targets for the loyalties of citizens, which ought rightly (on Hobbes’s view) to be directed at the sovereign. Hobbes holds, moreover, that persons are in general more likely to recognize and act on these more immediate loyalties than on their more distant obligations toward the state. He writes that “the common people have been, and always will be, ignorant of their duty to the public, as never meditating anything but their particular interests; in other things following their immediate leaders … as common soldiers for the most part follow their immediate captains, if they like them.”57 Associations can draw citizens’ attention away from their responsibilities to the political society in favor of obligations to their respective private groups.

Voluntary associations are additionally posed to undermine the stability of a regime insofar as participation in associations tends to draw citizens out of the relatively contained sphere of private life into the public arena. As Richard Boyd writes, the congregation of persons into purposive associations “delivers uncivil and disorderly individuals into political involvement en masse, carried along by the collective fervor of the ‘multitude’ or group.”58 When citizens are engaged in their private domestic or even commercial activities, the effects of their actions are unlikely to transcend the level of the quotidian and pose any challenge to the overarching framework of sovereign authority. In association, however, persons are encouraged to assume a public presence that amplifies the potential for conflict with political order.

56 Hobbes, Leviathan, chap. 29, par. 20, p. 218.
58 Boyd, Uncivil Society, 64.
The potential of associations to provoke conflict among citizens and to aggregate citizens into factions that could directly or indirectly challenge sovereign authority motivates Hobbes’s clear desire to strictly regulate private associations. Civil associations are to be permitted only at the grant of the sovereign, and their permissibility is to be judged with regard to their respective ends, and whether these are consistent with the maintenance of public order, “according to the lawfulness or unlawfulness of every particular man’s design therein.”59 In general, Hobbes inclines toward distrust of associations, affirming that “leagues of the subjects of one and the same commonwealth, where every one may obtain his right by means of the sovereign power, are unnecessary to the maintaining of peace and justice, and (in case the design of them be evil or unknown to the commonwealth) unlawful.”60 It is therefore the sovereign’s responsibility to regulate and, if necessary, suppress private association in order to minimize their threat to public order.

Hobbes recognized that pluralism poses a dual challenge to stability. Disagreement among the judgments, appetites, and beliefs of citizens provides the seed of conflict – either by frustrating attempts to resolve conflicts peacefully or by offending individuals’ sense of glory. Groups of citizens inflame individuals’ passions and thrust them into the public sphere, and also create objects of loyalty that may come into competition with the sovereign. Both of these challenges persist within civil society; they are not dispatched with the formation of the commonwealth. The maintenance of a stable regime must therefore be able to overcome or to mitigate the threats posed by pluralism.

60 Ibid., chap. 22, par. 29, p. 153.
Disagreement: Only a Byproduct of Freedom?

Rawls attributes the shift in his theory between *Theory of Justice* and *Political Liberalism* to his realization that the laws and institutions of a free society would inevitably give rise to a pluralism of comprehensive doctrines. This pluralism would undermine the congruence argument for stability that he offered in *Theory*. The fruits of a well-ordered society include a pluralism that required a new argument for the stability of that society. In my argument here, I have taken a robust pluralism as a given fact, not contingent on any particular set of political arrangements. But Rawls’s indication that pluralism is a product particularly of liberal societies suggests that maybe pluralism does not need to pose such a challenge for stability. Specifically, if pluralism and its attending challenges can be quashed by illiberal political arrangements, then perhaps the surest route to stability is by precluding disagreement altogether. If we are concerned only with stability and not with other goods such as freedom, why not simply suppress disagreement and eliminate the problems of pluralism? As we noted above, Hobbes’s solution is a version of this approach: prohibit the expression and teaching of those views which are contrary to the maintenance of sovereign order. (More on the scope of Hobbesian toleration below.) Why not go further and simply rule that only a single approved comprehensive doctrine can be approved or taught?

I may be able to beg off answering this challenge by recurring to my specification of this as a project in liberal theory. My attempt here is to contribute to our understanding of liberal political orders, so I am assuming that the societies in consideration will reflect some degree of freedom. This assumption is, in fact, what makes the project interesting. It is not much of a surprise that suppression of dissenting doctrines is one way to maintain stability. The real puzzle

---

is how a regime might achieve stability *despite* governing the interactions of persons who affirm a pluralism of comprehensive doctrines.\(^{63}\) At the very least, this is the aim of the present project.

There is also an empirical reason to doubt the suggestion that pluralism is purely a contingent feature of liberal political societies. The context in which contemporary societies exist makes the suppression of pluralism a practical impossibility. Even if such suppression were possible within a closed society, the increasingly pervasive reach of communication technology means that the flow of ideas from outside is difficult – if not impossible – to resist. The democratization of communication and media means that even persons living under autocratic regimes have access to a rich array of political and comprehensive views, exchanged in largely free and open fora.\(^{64}\) There are perhaps a few exceptions in the present day (North Korea being the most likely candidate), but the events of the so-called Arab Spring – and, importantly, the role of Twitter and electronic media in provoking those uprisings – have put on notice every other regime that attempts to maintain stability by imposing uniformity of belief. As the real world moves even further in this direction, Hobbesian confidence in the ability to suppress the emergence of pluralism will necessarily be diminished, and disagreement will become a feature of every society. A strategy of maintaining stability by means of suppressing disagreement, whatever its merits in theory, is simply not a viable option for any regime in the real world.

### 3.4 Impartiality in Hobbes

In light of the challenges that pluralism can cause for the stability of a political order, Hobbes endorses at least four principles reflecting some form of impartiality. For each of these,

---

\(^{63}\) Samuel Freeman helpfully pointed this out.

Hobbes takes the view that impartiality is required in order for the regime to win the support of those living under it and for destabilizing conflicts to be minimized. The four impartialist tenets of Hobbes’s political theory are: the rule of law, equality under the sovereign, toleration of religious difference, and the authority of public reason.

*The Rule of Law*

For those who associate Hobbes principally with the defense of monarchy and absolutist rule by a single sovereign, invoking his name in connection with a notion of the rule of law may seem absurd. Nevertheless, a number of interpreters, including Michael Oakeshott and Michael Zuckert, have emphasized the degree to which Hobbes’s theory is focused on the authority of systems of law, as opposed to the mere whim of an individual sovereign.65 These interpretations face the task of overcoming Hobbes’s apparent disdain for the notion of a sovereign law, captured in what has been called Hobbes’s “regress” argument, which appears to assert that stable order can be secured only by the rule of will – namely, the will of a single sovereign person.66 The regress argument is summarized most clearly in a passage from *Leviathan* where Hobbes rejects the notion that the sovereign himself must be subject to the constraint of law:

“Which error, because it setteth the laws above the sovereign, setteth also a judge above him, and a power to punish him, which is to make a new sovereign; and again for the same reason a third,

---


to punish the second; and so continually without end, to the confusion and dissolution of the commonwealth.” In other words, if the sovereign is subject to a law, then there must be some judge with the authority to punish the sovereign. That judge is therefore the true sovereign – but there must be someone with authority to punish that sovereign. Thus begins the regress that ultimately undermines sovereign authority altogether. On the basis of this argument, it would seem that Hobbes cannot endorse any notion of the rule of law.

This may be the view to which Hobbes is expressly committed, but a deeper look at his view suggests that the rule of law may be alive and well in the Hobbesian commonwealth. For one thing, as Jean Hampton has pointed out, a Hobbesian commonwealth must be governed by at least one rule, what H. L. A Hart called a rule of recognition. Specifically, a Hobbesian rule of recognition must identify as law whatever arises from an act of sovereign will. This is a trivial amendment, however. The more substantive way in which Hobbes’s commonwealth is governed by law concerns the formal restrictions Hobbes imposes on the will of the sovereign. While the sovereign has near-absolute latitude to determine the content of the law, the nature and purpose of sovereign authority requires that it establish uniform rule over its subjects. After all, a principal cause of the disorder that the sovereign is designed to mitigate is the conflict of dissenting private wills. The solution, then, must take the form of a singular authoritative public will. The content of that will, as expressed to the subjects, constitutes the civil law. Hobbes’s tenth through nineteenth laws of nature supply political norms which must be observed if legal order is to be stable, such as equity in dealing with subjects, equal use of common goods, and the

---

69 This is captured in Hobbes’s definition of civil law: “those rules which the sovereign hath commanded…” *Leviathan*, chap. 26, par. 3, p. 173.
necessity of impartial arbitrators.\textsuperscript{71} While Hobbes insists upon the undivided, unchallenged authority of the sovereign will, the purpose for which the sovereign is constituted requires that the sovereign rule by means of a law that is uniform and equitable.

\textit{Equality and Equal Terms of Cooperation}

It is well-known that in the state of nature persons are, by Hobbes’s account, naturally equal. Though many varieties of inequality must obtain among persons, they are equal in virtue of their vulnerability to violent death, which is the characteristic that most fundamentally defines the state of anarchy. Within the commonwealth, of course, a significant disruption is introduced into in the equality of persons: the sovereign occupies a position of absolute authority over each and every other member of the political society. But equality is not entirely discarded. Hobbes insists that all those subjects living under the rule of the sovereign must be regarded as equals.\textsuperscript{72} This is specified in his ninth Law of Nature, which reads “that every man acknowledge other for his equal by nature.”\textsuperscript{73}

Hobbes grounds this law in the natural equality of persons, but notes that it would hold even if it happens that persons are not actually equal by nature, because persons nevertheless regard themselves as equals. Hobbes’s claim here reflects his judgment that persons’ regular insistence on their own superiority\textsuperscript{74} or claim to rule\textsuperscript{75} evinces a psychology of equality: if each person thinks himself as good as or better than every other person, then none will accept terms of cooperation inferior to those enjoyed by others. Hobbes writes that “because men that think themselves equal will not enter into conditions of peace but upon equal terms, such equality must

\textsuperscript{73} Hobbes, \textit{Leviathan}, chap. 15, par. 21, p. 97.
\textsuperscript{74} Ibid., chap. 13, par. 2, p. 75.
\textsuperscript{75} Ibid., chap. 15, par. 21, p. 96.
be admitted.” Hobbes explicates the meaning of “equal terms” in his tenth law of nature, which specifies that “at the entrance into conditions of peace, no man require to reserve to himself any right which he is not content should be reserved to every one of the rest.” In other words, each party must join civil society in symmetrical manner, surrendering identical powers and reserving identical liberties.

Hobbes’s argument here is not immediately convincing. For one, the empirical claim that persons regard themselves as equals seems unfounded: persons routinely judge themselves as inferior to others in some manner or another. More basically, given the nastiness of the state of nature, why should persons refuse to cooperate, even if the terms are uneven? Is not unequal status in the commonwealth preferable to the chaos of anarchy? Rawls’s Difference Principle is premised on the notion that persons will rationally choose inequality if the well-being of the worst-off is raised in absolute terms. Surely even the worst off in the commonwealth are better off than they would be in the state of nature. So why accept Hobbes’s assertion that rational contractors will agree to enter political society only on equal terms?

The answer is glory. Hobbes takes it as axiomatic that persons have a psychological sensitivity to affronts to their own status. A person who would propose some superior status for himself by doing so offends against the desire for glory in each other person. (Hobbes calls this “arrogance;” the failure to acknowledge equality is “pride.”) Following the arguments we considered above, Hobbes views glory as a principal cause of conflict among persons. In order to

---

76 Ibid., chap. 15, par. 21, p. 97.
77 Ibid., chap. 15, par. 22, p. 97.
78 This thesis is also contained in Hobbes’s second law of nature, which prescribes that one ought “to lay down this right to all things, and be contented with so much liberty against other men, as he would allow other men against himself.”
overcome this problem in the state of nature, citizens must enter into civil order on equal terms; in order to prevent glory from undermining civil order once established, the law must continue to regard citizens as equals. This need not, of course, entail strict uniformity of treatment, for there are manners of differentiated treatment that do not imply inequality. (We considered this possibility in chapter 1.4.) The conclusion we can infer from Hobbes’s argument is that any differences in status must be such that they do not offend against persons’ desire for glory.82

Toleration

Though it remains a minority current within Hobbes interpretation, an increasing number of authors have devoted attention to the question of religion in Hobbes’s thought – an apt move, given that concerns of religion occupy most of the second half of *Leviathan*. The traditional reading of Hobbes on religion within the commonwealth emphasizes his Erastian commitments. Hobbes insists that the sovereign (who is, after all, a “Mortal God”83) retains the right to determine the canon of scripture, to specify the authoritative interpretation of those teachings, and to enforce religious teachings by legal means.84 More basically, Hobbes reserves to the sovereign the prerogative to suppress those doctrines – including religious ones – that are judged to be disruptive to the maintenance of civil peace:

“It is annexed to the sovereignty to be judge of what opinions and doctrines are averse, and what conducing to peace; and consequently, on what occasions, how far, and what

---

82 This might explain how the inequalities permitted by Rawls’s two principles of justice could accord with Hobbes’s tenth law of nature: so long as offices are open to all, persons who do not attain higher offices suffer no (necessary) offense.
men are to be trusted withal, in speaking to multitudes of people, and who shall examine the doctrines of all books before they be published.”

For those readers weaned on toleration in the Lockean vein, Hobbes’s stance might appear the complete opposite of liberal toleration of religion. In the last three decades, though, authors such as Alan Ryan, Richard Tuck, and others have drawn attention to evidence of “a more tolerant Hobbes.” Some of these readings emphasize the instrumental character of Hobbes’s endorsement of doctrinal suppression, and note that this account leaves wide latitude for toleration for religious views that do not (in the sovereign’s judgment) pose any threat to public order. Others go further, suggesting that Hobbes did not merely admit the possibility of toleration, but rather affirmed the good of independent conscience and free thought. In a late chapter in *Leviathan*, Hobbes writes that “there ought to be no power over the consciences of men but of the Word itself, working faith in every one … because it is unreasonable (in them who teach there is such danger in every little error) to require of a man endued with reason of his own, to follow the reason of any other man, or of the most voices of many other men…” These interpretations suggest a norm of religious toleration within the Hobbesian commonwealth, if bounded by the demands of order. I will not here wade into the scholarly debate over how robust

---

88 E.g., Curley, “Hobbes and the Cause of Religious Toleration.”
is this norm – or its bounds – but will stop at noting that within some ambit consistent with civil peace, Hobbes envisions a legal order that tolerates differing religious beliefs. That is, it treats them with some degree of impartiality.

Public Reason

I have mentioned a few times already the concept of public reason in Hobbesian theory. The idea of an authoritative public reason to supplant the private reasoning of citizens arises as a response to the problem of disagreement. As we discussed above, conflicts between judgment mean that even once the basic coordination problem of the state of nature is resolved, the risk of war still looms on account of disagreement. Hobbes is clear that these conflicting private judgments are insufficient for the determination of political things: “And when men that think themselves wiser than all others clamour and demand right reason for judge, yet seek no more but that things should be determined by no other men’s reason but their own, it is as intolerable in the society of men as it is in play…”90 The sovereign must, then, establish an authoritative source of judgments to which citizens agree to defer. This function is fundamental to the purpose of the sovereign, with Hobbes invoking it in his pivotal description of the formation of the commonwealth. In entering the commonwealth, persons agree to “confer all their power and strength upon one man, or one assembly of men, that may reduce all their wills, by plurality of voices unto one will, … and therein to submit their wills, everyone to his will, and their judgments, to his judgment.”91

The sovereign is not merely a common power, but a common judge. Hobbes notes this device as a familiar feature of conflict adjudication: “when there is a controversy in an account,

90 Ibid., chap. 5, par. 3, p. 23.
91 Ibid., chap. 17, par. 13, pg. 109 emphasis added.
the parties must by their own accord set up for right reason the reason some arbitrator to whose sentence they will both stand…”  

92 The sovereign performs this function writ-large in order to minimize conflicts arising from disagreement in the commonwealth. David Gauthier argues, though, that citizens do not simply abide by the judgment of the sovereign, but that they permit the sovereign’s reasoning to supplant their own. In order for peace to obtain, persons must acknowledge that their individual private judgments cannot be authoritative. But in order for each citizen to be able to identify with the sovereign’s “right Reason” – that is, for this common reason to serve its intended function – it must treat the concerns of each citizen equally. In other words, the sovereign must be impartial with respect to the private judgments of the citizens “in whose name” it exercises a public form of reason.  

Some debate has arisen over the scope and nature of the public reason that a Hobbesian theory requires. Gauthier, first and influentially, argued that Hobbesian public reason entailed transmuting the private judgment of one (the sovereign) into the public will of all. Gauthier held that Hobbes’s theory provides grounds to constrain the will of the sovereign, limiting his authority to those matters “that significantly affect the interactions of the citizens and the public goods available to them.”  

94 Michael Ridge has pointed out, however, that this solution does not resolve the problem of conflict, insofar as citizens’ judgments are bound to differ in regard to whether a particular issue falls within the sovereign’s jurisdiction, thus reigniting conflict among competing private judgments. In place of an authoritative “public person,” Ridge suggests that a Hobbesian solution could take the form of a set of principles of public reason, authorized directly by the subjects themselves.  

95 Ridge believes that his view avoids a problem of Gauthier’s,

---

92 Ibid., chap. 5, par. 3, p. 23.
93 Gauthier, “Public Reason,” 32–33.
94 Ibid., 37.
namely: if the sovereign’s authoritative judgment is only substantively constrained, what
guarantee do subjects have that his judgments will be equitable (as the tenth law of nature
requires)? Ridge’s prioritization of rules over persons “makes the standards of public reason
primary and derives when and how arbitrators should be authorized from those standards.”

Ridge does not specify the precise content of those rules, but presumably they would be
given by the laws of nature that Hobbes has identified, such as equity. Much more could be said
about the rules governing Hobbesian public reason – and I will return to the topic in the next
section. For now, the point is merely that Hobbes recognizes that stability requires some
common public source of judgment that citizens regard as authoritative, and that this public
judgment must be independent from the private reason of any particular citizen.

Each of the above sub-sections has focused on some element of Hobbes’s thought that
reflects a commitment to impartiality. How do these Hobbesian norms of impartiality relate to
the two theses – Distributive Impartiality (DI) and Justificatory Impartiality (JI) – that we have
considered in previous chapters? Recall that the theses read as follows:

*Distributive Impartiality:* Political arrangements ought to distribute benefits and burdens
without reference to politically irrelevant differences among persons.

*Justificatory Impartiality:* Coercive political arrangements ought to be supported by a
justification that is impartial with regard to comprehensive doctrines.

96 Ibid., 561.
97 Shane Courtland criticizes Ridge’s proposal on grounds that there is no plausible set of rules that is sufficiently
determinate as to avoid generating interpretive conflict. This is a compelling objection, but it is not fatal: a
satisfactory set of rules remains possible, if difficult to specify. “Public Reason and the Hobbesian Dilemma,”
Hobbes’s affirmation of the rule of law establishes the presumption of uniform treatment, with the corollary that any deviations from uniform treatment must be justified by reference to some “politically relevant” difference. In order for the law to distribute benefits and burdens differently to persons A and B, there must be some legally significant status difference between the two that grounds differential treatment. Establishing the rule of law does not tell us which differences are relevant (and therefore merit differential treatment), but it establishes the formal norm of Distributive Impartiality.98

The second and third impartialist tenets – equal terms and religious toleration – go some way to supplying content to Hobbesian DI. The requirement of equal terms implies that whatever inequalities exist in the state of nature (and therefore might be invoked to justify unequal terms) are insufficient to ground differential treatment. So physical strength, intelligence, and so on, are taken to be politically irrelevant. Hobbes also specifies that wealth is not a relevant grounds for differential treatment.99 These are not positive specifications of political relevance, but rather exclude several categories of difference from consideration in the distribution of benefits and burdens. The norm of toleration excludes an additional category: religious belief. Differences in belief is not itself a legitimate grounds for differential treatment. The seditious nature of some beliefs does indeed constitute such a grounds, but differences in religious identity does not. Thus, we can say that the Hobbesian sovereign must be impartial toward his subjects with respect to (at least) strength, wit, wealth, and religious identity.

99 “Seeing, then, the benefit that everyone receiveth thereby is the enjoyment of life, which is equally dear to poor and rich, the debt which a poor man oweth them that defend his life is the same which a rich man oweth for the defence of his…” Leviathan, chap. 30, par. 17, p. 227.
The notion of a Hobbesian public reason relates clearly to the thesis of Justificatory Impartiality. The function of the sovereign’s public reason, which supplants the private reasons of the individual subjects, is to make determinations of right and wrong and, thereby, to determine the content of the law. Law is thus justified by the judgment of the sovereign. For Hobbesian public reason to serve its conflict-mitigating function, it must be independent from the subjects’ private judgments. Hobbes does not focus specifically on “comprehensive doctrines” or full-blown religious/metaphysical/ethical systems of belief, but the basic requirement of Hobbesian public justification nevertheless tracks JI closely: laws must not rely on the judgment of any individual subject for their justification, but rather must have some (public) justification that is impartial with respect to the truth value of subjects’ private reasons. The reason of the sovereign, in other words, must be impartial with respect to subjects’ individual beliefs.

We see, therefore, that the rule of Hobbes’s sovereign reflects both distributive and justificatory impartiality. The task of this section has been to show that this is true as a matter of exegetical fact. Hobbes thus provides helpful resources for a stability-based account of impartiality, and the fact that the theorist of stability par excellence affirms these norms of impartiality counts as presumptive support for my claim that stability does indeed require impartiality. However, the details of Hobbes’s arguments are not always clearly stated, and thus the relationship between stability and impartiality is uncertainly established. In the following section, we turn to the task of connecting the logical dots more clearly, constructing a positive account of how the end of stability motivates the two theses of impartiality.
3.5 Stability and Distributive Impartiality

As we did in 1.4 for the account of impartiality beginning from equality, the objective of this section and the next is to articulate clearly an account of impartiality that begins from stability. Hobbes has served as our guide thus far, both in highlighting the threats to stability that arise from pluralism and in affirming several manifestations of impartiality. In the following, we will depart from a focus on Thomas Hobbes the person and will shift into a constructive philosophical mode that is Hobbesian only in the sense that it takes the concerns and the insights of Hobbes seriously. While I suspect that Hobbes would affirm much, if not all, of what follows, I want to be clear that my intention is not to attribute the following argument to Hobbes himself.

One clear way in which we will depart from Hobbes is by explicitly adopting stability for the right reasons as the foundational end from which to derive principles of impartiality. As discussed above, Hobbes’s conception of the necessary conditions of stability is not as simplistically reliant on coercion as the caricature of his view might suggest, but stability for the right reasons goes further, beginning with a consideration of the necessary conditions for a regime to win the non-coerced endorsement of citizens, and invoking coercion only as a non-ideal, last-resort resource. To take SFRR as our starting point is to ask what kind of regime can win the support of those who live under it – by means other than coercive force. Our task is to identify the ground of each citizen’s sufficient reason to maintain a commitment to compliance with the regime.

A second way in which we will part ways with Hobbes is by dropping the language of “the sovereign” in favor of a less determinate description of regime. Hobbes is clear that while rule by an individual person is not strictly necessary for stability, it is most certainly to be
preferred. For the moment, I wish to set aside the specific question of regime-type, neither assuming that stability is achievable only by the rule of an absolute monarch nor following too quickly those who have asserted that Hobbes’s own argument is fully compatible with separation of powers, democratic control, or other non-absolutist forms of governance. For all intents and purposes, the argument which follows is agnostic on the form of regime most conducive to stability.

One final caveat before proceeding to the argument. I do not here intend to offer a comprehensive and exhaustive account of the political principles, institutions, practices, and so on that are required for stability (for the right reasons). Such a project is beyond my ability here, and also beyond the scope of the present project, which is focused narrowly on the question of pluralism and accommodation of difference. In light of this focus, my objective is to establish the logical relationship between stability and impartiality – both distributive (in this section) and justificatory (in the next). These are, I will argue, two necessary features of a regime that can become stable for the right reasons. There may be other features that such a regime must possess, but I will not attempt to identify all of those here.

With these preliminaries established, we may proceed to the argument. The first step is to stipulate that for the achievement of the kind of stability we are interested in (stability for the right reasons), the regime must generate its own support by supplying citizens with reasons that can win reflective endorsement to maintain their support. The definition of SFRR that we formulated in 2.1 specified that the support generated by the regime must be sufficient to secure the likely endurance of the regime, which does not necessarily imply that each and every citizen recognizes a sufficient reason for such support. Rather, only a sufficient number of citizens must do so. For present purposes, I leave the specification of this number unspecified, with the

\[100\text{Ibid., chap. 22, p. 146–55.}\]
understanding that it refers to whatever portion of citizens whose support is sufficient to ensure the likely endurance of the regime.

(1) SFRR is possible if and only if most citizens recognize a sufficient reason to maintain a commitment to supporting the regime.

One thing that will help us is clarifying what constitutes a “sufficient reason.” I take a sufficient reason to be one that is not overridden by some other reason for contradictory action. (Raz calls this a “conclusive” reason.101) Thus, one way to show that a person does not have a sufficient reason to do some action is to show that there is some other stronger reason to do a different action. So one way to show that impartiality is required for stability is to show how partiality would give citizens reasons not to support the regime that would override the reasons they recognize to support the regime. This is roughly the structure of the Hobbesian arguments for impartiality that we considered above. There are at least three lines of argument we can draw from Hobbes to ground the thesis of DI in the end of stability for the right reasons.

The first argument suggests that, because conflict is inevitable, people need recourse to an independent, impartial arbiter. The reasons for conflict may be any of those we considered above, including Hobbes’s classic three sources – competition, diffidence, and glory – along with the more complex sources of conflict engendered by pluralism. Regardless of the origin of conflict, we need an impartial third party to whom we can appeal for resolution. The necessity of some sort of judge to ensure resolutions determined other than by contests of raw power is obvious. (Along with Hobbes, John Locke is of course the other major figure to situate the need

for a third-party arbiter centrally in his justification for the constitution of political authority.) Some conflicts, of course, are most appropriately worked out between the involved parties, but unconstrained conflict is incompatible with stability for the right reasons.\textsuperscript{102} Stability thus requires an arbitrator.

Why, though, must this judge be impartial? Would it not be rational for persons to prefer \textit{any} resolution to continued conflict, regardless of whether the arbiter were to produce such a resolution impartially? The short answer is no. Conflict presupposes that each person judges their own position, interest, or preference to have some legitimate claim to triumph over that of their opponent. Each person believes in the rightness of their side. A judge who does not rule impartially tilts the balance against one or the other party, independent of the merits of their claim. Imagine, for instance, a judge who rules always in favor of the wealthier party. Very poor persons would have only a very weak reason to submit to such a judge, for in most cases they would be pre-determined to lose. Even people of moderate wealth might prefer to do battle on the merits of their claims, rather than risk going before a judge as the poorer party in the dispute.

Of course, a judge could avoid partiality by not acting on any reliable preferences, but rather making decisions in an unpredictably capricious way. This is impartiality of a sort, though only trivially.\textsuperscript{103} But it won’t suffice to win support either. If my odds entering into arbitration are no

---

\textsuperscript{102} For the sake of clarification, unbounded conflict is tantamount to instability because SFRR assumes a system of social cooperation. If a social/political situation does not reflect organized cooperation, then it is a modus vivendi at best, as persons are committed not to the maintenance of some standing, mutually-benefitting order, but rather to their own individual advantage. Even once political order is established, conflict unconstrained by that order produces instability by virtue of the erosion of social cooperation, which is a precondition for SFRR. Hobbes adopted a cautious approach to any conflict whatsoever, worrying that even small disputes would tend to magnify into politically disruptive antagonism. The warning is well-taken, though it neglects the role that political systems can play in bounding conflicts that might otherwise escalate, and thereby potentially \textit{facilitating} conflict of a constrained sort. My neighbor and I can argue with relative confidence that one of us won’t kill the other because murder is against the law and carries stiff penalties. Absent that legal framework, we may be more reluctant to confront one another in the first place, for fear of where our conflict might lead. This being said, the basic claim that conflict tends to undermine stability stands.

\textsuperscript{103} I called this a trivial form of impartiality in the Introduction.
better than a coin toss, then I will rationally prefer to remain in conflict in hopes of my position (which I believe is superior to my adversary’s) coming out on top.

In short, the inevitability of conflict means that some party must be empowered to serve as an arbitrating judge. In order for an arbitration system to win support from those whose conflict it is meant to adjudicate, the judging mechanism must be impartial toward the parties to those conflicts. Otherwise, persons will have overriding reasons to reject the arbitration system and remain in conflict. We might call this the Impartial Arbiter argument. It can be captured, along with the inevitability of conflict, as follows:

1. Conflict among persons is inevitable both prior to the constitution of political order and within it.

2. Conflict is incompatible with stability.

3. The resolution of conflict requires that parties submit to the judgment of an arbiter who is impartial toward them.

The second line of argument follows Hobbes’s ninth law of nature, which asserts that persons entering into a system of social cooperation will only agree to do so if the terms of participation are equal – that is, if each party agrees to the same forfeits of power in exchange for the same benefits. Hobbes thinks that this is the case because persons regard themselves as equals, so therefore they will only accept cooperation on equal terms. This premise, as indicated above, is descriptively questionable. Persons do not without exception think of themselves as
equal to one another, even in the most important respects. But even if they did, it would not be rational to decline the protections of the state of nature on grounds that the terms of cooperation fell short of perfect equality. The problem of glory, though, does provide some support for Hobbes’s claim. Because people are sensitive to perceived slights, they will choose conflict over peace if they judge the terms of peace offensive to the status they think commensurate to the esteem they hold for themselves. That a state of affairs involves accepting an insult to one’s sense of glory is a reason to reject that state of affairs that might override a reason to accept that state of affairs based on whatever other benefits it may provide.

This claim is compatible with our skepticism about persons’ insistence on their status as equals. Even if persons think of themselves as unequal, there will be some standard of treatment that each believes herself to be due in keeping with whatever she believes her status to be. Even if each person would not insist on equal terms, profoundly undignifying terms might still offend some persons’ sense of glory. For instance, imagine a committed Aristotelian who is fully cognizant of his own inferior virtue, and who might therefore accept some inferior political office compared with more excellent compatriots. This person could nevertheless maintain that he is not entirely worthless, and could therefore judge political arrangements that treat him like dirt as offensive to his status.

The riots that occurred in Baltimore during May of 2015, according to some interpretations, point up the instability of a regime whose distribution of burdens and benefits does not track citizens’ subjective estimation of their status. Rioters explained their actions as a response to long-standing disadvantage, attaching primarily to race, though also geography. The violent and destructive outburst had to do not only with the episode of police brutality that served as an immediate catalyst, but with a distrust of the political order that had simmered for many
years. Political arrangements that subject persons to disadvantage incommensurate with the equal status they estimate for themselves give rise, over time, to unrest and resistance. The Baltimore riots did not seriously threaten to upset the political order of Baltimore, Maryland, or the United States, but they hinted at the threat to stability that can be generated by offenses to citizens’ subjective sense of worth.

The significance of the human desire for glory in achieving stability, then, is that in order for persons to endorse the political arrangements under which they live, those arrangements should track as closely as possible those persons’ estimations of themselves and the recognition they are due. There are reasons to think that equal terms of cooperation may best achieve this end. For one, Hobbes is correct to note, of course, that persons’ estimations of their own status will often conflict: not everyone can be the best, and outside of Lake Wobegon, not all children can be above average. Equal terms at least avoids any person taking offense at having been granted an inferior status to someone else. Moreover, persons’ estimation of what they are due will inevitably be shaped by contingent facts about social context (along with individual psychology, etc.), and in the context of Western democracies, the presumption of equal regard is deeply engrained. Differences in individuals’ subjective estimation of the status each is due may therefore be muted in democratic contexts.

For these reasons, establishing equal terms of cooperation is likely the most effective way to minimize slights to glory – and therefore to tilt citizens’ balance of reasons in favor of endorsing the cooperative pact. This is a contingent conclusion, however: the terms of cooperation would be subject to modification in light of persons’ actual expectations. The goal will be to find the system of cooperation that is Pareto-optimal in terms of minimizing perceived offense. The argument from glory establishes a presumption of equal terms. Any deviation from
this norm bears the justificatory burden of showing how asymmetric terms of cooperation would better avoid destabilizing offenses to citizens’ desires for glory. Formulated expressly in terms of impartiality, the norm of equal terms entails that differences among persons are presumed to be insufficient grounds for a differential political distribution of burdens and benefits. Call this the *Presumption of Equal Terms* argument:

(5) Persons are sensitive to perceived offenses by political arrangements that do not accord them the status they feel is due in light of their particular characteristics.

(6) A norm of distribution of benefits and burdens that does not take into account differences among persons will effectively minimize offense in most cases. Deviations from this norm may be warranted, but if and only if such deviations more effectively minimize offense.

A third Hobbesian argument is implicit in Hobbes’s first law of nature, and is aimed directly at one of the fundamental coordination problems that a stable political system must overcome: mutual assurance. It is, of course, necessary for peace that each person agree to lay down his arms and enter into cooperative association. However, one can – indeed, *ought* – do so only if one has assurance that others will do the same. Hobbes’s fundamental law of nature specifies that “a man be willing *when others are so too* … to lay down this right to all things…”\(^{104}\) When moving from the state of nature to the commonwealth, the necessity of this assurance is obvious, as without it no person would judge it rational to adopt a cooperative stance. The need for assurance persists even within civil society, since a person who comes to

doubt that her fellow citizens are likely to comply with the requirements of the regime, will become less likely to do so herself.

One way a sovereign authority can provide assurance that parties’ compliant acts will be met by reciprocal compliance is by imposing coercive penalties on acts of non-compliance, such that each party can have confidence that every other’s balance of reasons tilts, by virtue of those penalties, in favor of compliance. But it is part of the definition of SFRR that citizens’ reasons for supporting their regime cannot be grounded in coercion. If the support that makes the endurance of the regime possible is derived only from coercion, then it is not stability for the right reasons. There must be other, non-coercive reasons deriving from the character of the regime for citizens to maintain their commitment to support it. But these are less clearly observable by third parties: I have no way to tell if (for instance) some other citizen’s moral beliefs affirm the political principles around which the regime is structured. SFRR therefore relies on sources of support that are not readily apparent from person to person.

This is the case even when I observe my fellow citizens complying with the requirements of the regime. While I may appreciate their compliance, I do not know if they are interested primarily in advancing their own interest – in which case they view the present order as a mere modus vivendi – or if they are committed to supporting the reciprocal system of social cooperation. If I suspect that my fellow citizens view our arrangements as a modus vivendi, then it is rational for me to do the same, eyeing each and any opportunity to achieve more favorable terms of cooperation. Thus, robust stability requires that I have confidence not only that others will comply with the terms of our political arrangements, but that they are committed to upholding those arrangements – not preparing to jettison them at the first chance to secure something more individually advantageous.
It is impossible to have perfect confidence of this sort, as one can never fully know the beliefs and intentions of every of one’s fellow citizens. But it is easier to have this confidence when the terms of cooperation are symmetrical. Here we arrive back at the norm of distributive impartiality. When political arrangements afford differential treatment among persons or groups thereof, persons may reasonably grow wary that their fellows (particularly those enjoying uncommon benefit from the regime) support the regime on only a contingent basis. In contrast, when benefits and burdens are distributed without preference for this or that class of persons, I can recognize that the dynamics of commitment are identical for me and for others. On the basis of this recognition I can form a belief that my fellows are committed to the regime in a spirit of reciprocity – and thereby I can judge it rational for me to maintain my commitment to support the regime.

Consider, by way of example, a regime that abided by a distributive principle that specified special benefits for those citizens belonging to a particular religious tradition. Persons of that religious identity would of course find reason to support the regime: in addition to whatever benefits the regime provided to all citizens, they would be able to count on the additional perks attaching to their particular identity. Others, however, would have reason to doubt the character of their commitment. Are members of religion X committed to an effort in social cooperation, or do they support the regime merely because it provides them with special benefits? If those special benefits were to be lost, and members of religion X were to be treated like every other citizen, would they maintain their support? If members of religion X are committed primarily to their own individual or group advantage, then their commitment is contingent on the particular balance of forces that has resulted in their preferred status. Other citizens (particularly those who receive special burden) will reasonably adopt a suspicious eye
toward members of religion X and will adopt a similarly self-interested attitude toward the regime, waiting for a moment when the balance of powers shift such that they can reallocate those special benefits to their particular group.

This may explain why public officials who dodge taxes or shelter their money in offshore accounts tend to meet with public scorn. Their efforts to evade the burdens of membership in political society, even when they are legal, signal that they are committed principally to advancing their own financial well-being. This may suggest that their support of the regime is merely strategic, and that they would readily shift the terms of cooperation to benefit themselves further if given the opportunity. The same worry may even be magnified when connected to group identity. As Hobbes noted, private associations can attract loyalties potentially in conflict with citizens’ loyalty to the political order. More to the present point, they can give the appearance of conflicting loyalties, particularly when their members appear attached to the regime only because of the special benefits that flow to their preferred group. In short, differential treatment undermines mutual assurance of reciprocal commitment to supporting the regime and encourages citizens to adopt a mode of support consistent with a modus vivendi rather than stability for the right reasons.

Impartiality is therefore supportive of mutual assurance, insofar as it eliminates the cause for suspicion that persons are committed to the regime only in the manner of a modus vivendi. This is a different conclusion from those supported by the Impartial Arbiter and Presumption of Equal Terms arguments, which held that distributive impartiality is a necessary condition of stability for the right reasons. The Mutual Assurance argument, in contrast, affirms merely that distributive impartiality is supportive of SFRR, though not necessary for it. At the most, this supports a presumption in favor of DI, but not conclusive evidence for its necessity.
(7) SFRR requires that each citizen be reasonably assured of others’ likelihood of supporting the regime.

(8) Political arrangements which distribute benefits and burdens with regard to differences among persons provide grounds for suspicion that persons supporting the regime do so only as a modus vivendi.

(9) If one suspects that others support the regime only as a modus vivendi, it is rational for one to adopt a similarly contingent stance of support.

Each of the three arguments we have considered in this section provides independent support for the thesis of Distributive Impartiality. The Impartial Arbiter argument casts an impartial distribution of benefits and burdens as a functional requirement of the political order as an arbitrator of inevitable conflict between persons. The Presumption of Equal Terms argument holds that an impartial distribution will most effectively minimize offenses to citizens’ desires for glory. The Mutual Assurance argument suggests that impartial distribution supports citizens’ confidence in one another’s reciprocal commitment to supporting the regime. None of these fully gives content to the norm of DI. They do not tell us what counts as a “politically relevant” difference that would warrant differential treatment. But they support the general norm of impartiality and establish a burden of justification for deviations from it. The categories of difference that Hobbes excludes – strength, wealth, wit, religious identity – would likely be similarly excluded by the arguments we’ve advanced here. To achieve stability for the right
reasons in the face of persistent challenges arising from disagreement and difference, the political order ought to reflect a general norm of impartiality with respect to those differences.

I have said in a couple of places already that I would mostly leave aside the issue of impartiality in application of the law, but it merits noting here that the arguments for Distributive Impartiality in the content of the law would also support impartial application of those laws. Each of the arguments against differential treatment by the law can be extended to conclude that the laws must be applied impartially to have their desired effect. Hobbes recognized this expressly, noting the destabilizing consequences of partiality in enforcement of the law: “The safety of the people requireth further … that justice be equally administered to all degrees of people … so as the great may have no greater hope of impunity when they do violence, dishonour, or any injury to the meaner sort, then when one of these does the like to one of them.” If it is not, “The consequences of this partiality towards the great proceed in this manner. Impunity maketh insolence; insolence, hatred; and hatred, an endeavour to pull down all oppressing and contumelious greatness, though with the ruin of the commonwealth.”105 Hobbes here emphasizes the resentment that would develop if some persons were permitted to break the law while others were not. The issue of mutual assurance also supports the need for impartial application, insofar as partiality in this regard specifically undermines confidence that others will be likely to follow the law. In short, though impartiality in application of the law is distinct from laws that are impartial in their content, both kinds of impartiality are required for the stability of a political order.

105 Ibid., chap. 30, par. 15–16, p. 226–27.
3.6 Stability and Justificatory Impartiality

The requirement of distributive impartiality is a constraint on the content of laws for the sake of stability. It tells us relatively little about the array of laws and institutions that ought to populate a political order, other than that they should be in some respect impartial. But within these bounds, myriad political questions remain: What should the laws be? How should they be interpreted in cases of conflict? Working out these details is bound to be a contestatory enterprise. This much is obvious from ordinary experience, but it is also supported by the Hobbesian premises identifying the conflict-generating aspects of pluralism. Principally, in the specification of laws and political institutions, individuals’ judgments will inevitably clash. Persons will disagree about what kind of political arrangements are best. I have argued in section 2.4 that the normative ideal of stability for the right reasons does not preclude disagreement and political contestation. What is necessary for stability is to properly contain or accommodate the effects of disagreement such that they do not undermine citizens’ reasons to maintain their support for the regime.

Hobbes’s solution was for members of the commonwealth to subordinate their private judgments to that of the sovereign, at least in public matters.¹⁰⁶ This procedural solution transforms the private reasoning of one individual – the sovereign – into the public reason of all. Having each member of the commonwealth submit to the judgment of the sovereign is one way of satisfying the requirement of SFRR that each person recognize, from a perspective of reflective deliberation, some sufficient reason to endorse the regime. If I have made a prior decision that the reason of the sovereign is authoritative in matters public, then the fact that the sovereign endorses this or that law or institution is sufficient for me to endorse it as well.

¹⁰⁶ Recall Gauthier’s specification of the bounds of the authority of the sovereign’s reason as encompassing those matters “that significantly affect the interactions of the citizens and the public goods available to them.”
Liberals might reasonably worry that this goes too far, however. Absolute deference to the judgment of the sovereign raises concerns of “creating a monster.” But Hobbes’s solution is not the only way to meet the justificatory requirements of stability, however. Subsequent theories of public reason – from Locke through Rawls (and beyond) – have sought to reconcile the maintenance of individual belief and judgment with the stability of public reason. My aim here is to show that the norm of justificatory impartiality can achieve this end without requiring uniformity in citizens’ private judgment.

Recall that justificatory impartiality specifies that “Political arrangements ought to be supported by a justification that is impartial with regard to comprehensive doctrines.” For a law to be justificatory impartial means that “with respect to set of reasons S, there is no member M of S for which the fact that M is false implies that L is unjustified.” In other words, in order for justificatory impartiality to obtain, political arrangements must not rely for their justification on any particular comprehensive doctrine. As was the case for the norm of distributive impartiality, there are multiple lines of argument that ground justificatory impartiality on the basis of stability.

First, SFRR requires that citizens each recognize some sufficient reason to endorse the regime. (Recall that by “regime,” I mean both a set of political principles and the laws and institutions that instantiate them. So by “endorsing” a regime or a regime being “justified,” I refer to endorsement or justification of the principles on which the laws and institutions are structured.) This means that, at the very least, the justificatory reasons supporting the regime must not be basically opposed to any citizen’s comprehensive doctrine. If they were, that

---

109 See Introduction.
110 Rawls says that in an overlapping consensus the political conception of justice must “not be too much in conflict” with citizens’ essential interests. Rawls, Political Liberalism, 134.
citizen’s comprehensive views would override the justificatory reasons supporting the regime, potentially tilting the citizen’s balance of reasons against maintaining her support for the regime.\footnote{Gaus calls this “the problem of defeated endorsement,” which he distinguishes from the related problem of “ineffective endorsement.” For present purposes, I will ignore this distinction and treat the two problems together. Gerald Gaus, “A Tale of Two Sets: Public Reason in Equilibrium,” \textit{Public Affairs Quarterly} 25 (2011): 307–8.} Avoiding this kind of conflict might be achieved in at least a couple different ways, captured by the so-called “consensus” and “convergence” conceptions of public reason.\footnote{D’Agostino, \textit{Free Public Reason}, 30; Kevin Vallier, “Convergence and Consensus in Public Reason,” \textit{Public Affairs Quarterly} 25 (2011): 261–80.}

On the first, the regime could be justified on the basis of a single set of shared reasons that are consistent with everyone’s comprehensive views. Alternatively, the regime could be justified on the basis of various supporting reasons which are not shared by all citizens. So long as each citizen recognizes some such reason that is compatible with their comprehensive beliefs, then the problem of overriding comprehensive doctrines is avoided. In more precise terms, on the consensus model, there exists a set of reasons justifying the regime that is also a subset of each citizen’s set of reasons compatible with her comprehensive doctrine. On the convergence model, in contrast, there is no unique set of justifying reasons that belongs as a subset to each citizen’s set of reasons compatible with her comprehensive doctrine, yet each citizen nevertheless recognizes some non-shared reason to endorse the regime. There is no consensus on why the regime is justified, but each citizen nevertheless believes that it is. The regime is then capable of winning endorsement from the perspective of many different comprehensive views, even if the reasons motivating that endorsement differ.

There may be good reasons – both on the basis of stability and otherwise – to prefer one over the other, but I will not wade into that debate here.\footnote{In favor of convergence, see Gaus, “A Tale of Two Sets”; Gerald Gaus, “Hobbes’s Challenge to Public Reason Liberalism: Public Reason and Religious Convictions in \textit{Leviathan},” in \textit{Hobbes Today: Insights for the 21st Century}, ed. S. A Lloyd (Cambridge: Cambridge University Press, 2013), 155–77; John Thrasher and Kevin Vallier, 113} The point is that both of these models
of endorsement satisfy the definition of justificatory impartiality: in either case, the regime does not depend for its justification on the truth of any particular comprehensive doctrine. The only context in which this would not be the case is a scenario where citizens unanimously affirm a single comprehensive doctrine and, on the basis of that doctrine, endorse the regime. (In this case the distinction between consensus and convergence actually collapses.) A theocratic regime based on a religion with which the whole of society identifies would secure the endorsement necessary for stability. But the fact of pluralism rules out this possibility. We must assume that members of a society will affirm a variety of comprehensive doctrines, at least some of which will be incompatible with one another. From this, it follows that:

(10) For any comprehensive doctrine C affirmed by some citizen(s), there will be some other citizen(s) whose own comprehensive doctrine(s) reject the truth of C.\textsuperscript{114}

If this is the case, then there is no possible comprehensive doctrine that can win endorsement from the entire citizenry. The theocratic state can be stable only in a nonpluralistic context, which is to say, not at all.

(11) A regime that depends for its justification on some particular comprehensive doctrine C will fail to win endorsement from an existing set of citizens whose comprehensive doctrines reject the truth of C.

\textsuperscript{114} They may either affirm the falsehood of C expressly or may affirm some alternative comprehensive doctrine which implies the falsehood of C.

\textsuperscript{114} They may either affirm the falsehood of C expressly or may affirm some alternative comprehensive doctrine which implies the falsehood of C.
Stability, in short, requires that citizens recognize some reason to reflectively endorse the regime under which they live. By (10) and (11), in order to meet the end of stability, a regime must not rely upon any one comprehensive doctrine for its justification. In other words:

\[ (JI) \quad \text{Coercive political arrangements ought to be supported by a justification that is impartial with regard to comprehensive doctrines.} \]

This is the most straightforward line of argument supporting justificatory impartiality on the basis of stability. But there is another argument, drawing on the Hobbesian insights regarding glory and persons’ propensity to take offense at disagreement. That a set of political principles is instantiated in a system of laws and institutions, and moreover that persons endorse and support that system, constitutes a robust public affirmation of those principles and any anterior propositions on which the principles rely for justification. A stable political order reflects the convictions of the public it governs, and can be taken as an indication of the beliefs that those supporting the regime hold to be true.

\[ (12) \quad \text{A stable regime implies the truth of any propositions on whose truth the justification of the regime depends.} \]

With this in mind, consider the effect of a regime that relied for its justification on some particular comprehensive doctrine. Those citizens who affirm that doctrine would of course endorse the regime, but moreover might reasonably feel cheered at the fact of their system of
belief being implicitly affirmed by the political order.\textsuperscript{115} Citizens who do \textit{not} affirm that doctrine, in contrast, might reasonably take offense at the apparent public slight to their own system of belief. As Abizadeh notes in characterizing Hobbes’s view of the offense given by disagreement, “the simple expression of a different opinion is by far the most devastating sign of contempt.”\textsuperscript{116} Even more so when that expression is given not by another mere individual but by the state itself. Consider, for example, debates over the presence of monuments to Judeo-Christian religious beliefs (such as representations of the Ten Commandments) on public grounds. At issue centrally in these debates is whether such displays imply state endorsement of the corresponding religious beliefs. People who do not believe in the divine origin of the Ten Commandments may be reasonably offended at the public recognition of a competing religious system – and the implied dismissal of their own beliefs. These cases have ignited passionate responses on both sides, and yet they concern a relatively minor aspect of political life: the decoration of public buildings. In a 1984 Supreme Court case regarding a nativity display erected by a city in a public park, Justice Sandra Day O'Connor noted the danger inherent in the display's implicit endorsement of Christian faith: “Endorsement sends a message to nonadherents that they are outsiders, not full members of the political community, and an accompanying message to adherents that they are insiders, favored members of the political community.”\textsuperscript{117} Political arrangements – no matter how apparently inconsequential – that imply the affirmation of this or that religious belief system cause inherent offense to those who do not share in the belief. Citizens taking offense at the slight to their beliefs (and therefore to them as well) by the political order’s implied endorsement of a competing comprehensive doctrine may be tempted, as Hobbes notes, to respond

\textsuperscript{115} In case the logic of this implicit endorsement is not clear: that principle P is justified implies the truth of any premises that are necessary for its justification.

\textsuperscript{116} Abizadeh, “Hobbes on the Causes of War,” 308.

conflictually. Even if they do not go this far, they may reasonably experience a diminished allegiance to their regime.

(13) By (11) and (12), a regime that depends for its justification on a particular comprehensive doctrine will offend those citizens who reject that comprehensive doctrine, thus diminishing their support for the regime.

This offense may be avoided, however, when the regime does not presuppose the truth of any particular comprehensive doctrine. In other words, justificatory impartiality.

Thus, the end of stability supports the requirement of Justificatory Impartiality in two ways. First, in order for each citizen to recognize some sufficient reason to endorse (and support) a regime, that regime must not depend for its justification on any particular comprehensive doctrine. Second, the same must hold in order to avoid giving potentially destabilizing offense to those who would not affirm whichever comprehensive doctrine is implicitly affirmed by a justificatorily partial regime.

Note, though, that not every justificatorily impartial regime will be capable of achieving stability (for the right reasons). The reason is that there may be some impartial regimes that cannot be affirmed by each and every citizen. Consider a scenario in which a regime is affirmed by a third of the citizens on the basis of comprehensive doctrine A and by a second third of the citizens on the basis of some other comprehensive doctrine B (where A and B are entirely independent of one another), but which is rejected by the final third of citizens. This regime does not depend on any one comprehensive doctrine for its justification, for if either A or B were
false, it would remain justified on the basis of the other. But nevertheless, the regime would fail to provide one third of its citizens with sufficient reason to for endorsement.

What this shows is that justificatory impartiality is a necessary but insufficient condition for stability. This is no problem for our argument so far – the aim has been merely to show that JI can be derived from the end of stability. It is nevertheless important to clarify that the two theses of impartiality are not intended to describe exhaustively the requirements for stability. Rather, the arguments in this chapter have established that the normative ideal of stability for the right reasons, whatever else it requires, requires both distributive and justificatory impartiality. These can be justified on the basis of stability alone, without appeal to equality – the traditional basis of impartiality in liberal theory. The norms of liberal impartiality thus need not rest on substantive moral premises, but rather on the uncontroversial end of stability. Properly conceived (as I argued in Chapter 2), stability requires that a regime meet conditions for endorsement by those who live under it. This chapter established that two such conditions are distributive and justificatory impartiality. In the next chapter, we turn to a challenge facing impartiality – and therefore stability – that concerns the psychological conditions necessary to effectively motivate citizens to maintain support for a regime that is impartial with respect to them and their comprehensive doctrines.
CHAPTER FOUR:

OVERCOMING IMPARTIALITY’S MOTIVATIONAL DEFICIT: SOME APPROACHES

The project of this dissertation so far has been to defend two general principles of impartiality, one pertaining to distribution of benefits and burdens, the other pertaining to the normative justification of political arrangements. A regime is distributively impartial insofar as it distributes benefits and burdens without respect to certain features of the persons who live under it, such as their genders, racial identities, or religious beliefs. A regime is justificatorily impartial insofar as the laws and institutions that compose the regime do not depend for their justification on the truth of any one comprehensive doctrine. The principles of distributive and justificatory impartiality normatively describe a set of political norms familiar to students of the liberal tradition – and in particular contemporary accounts of political liberalism.

The argument of Chapter 3 was that a regime must satisfy the general principles of distributive and justificatory impartiality in order to win the endorsement – a necessary condition for support – of the citizens it governs. In this chapter, we consider a different question, namely: what can reliably motivate people to act in accordance with these norms? The reason for taking up this question is that rational endorsement does not always entail practical motivation. The rational justification of political principles may not translate directly to citizens behaving in ways that comply with and uphold these principles. I argue in this chapter that liberal impartiality faces exactly this challenge, and regimes must therefore look beyond the basic justification of impartiality to find sources of effective support. More specifically, a regime must draw on the objects of citizens’ affection, passion, and commitment as resources in order to compel not only rational endorsement but also practical support.
The core argument of this chapter is that religious belief can provide one such resource, and I draw on both John Locke’s “Letter Concerning Toleration” and John Rawls’s notion of the overlapping consensus as conceptual illustrations of how support for impartiality may be motivated by citizens’ distinctive religious commitments. Before considering these, however, I address two other potential approaches to the problem of motivating support for impartiality. One seeks to generate motivation by uniting citizens around a common civic identity; the other aims directly at citizens’ emotional faculties, working to elicit patriotic sentiments of support for the regime. I identify the former strategy in Rousseau’s social contract theory and in Jürgen Habermas’s idea of constitutional patriotism. The latter strategy has been recently advanced by Martha Nussbaum. I argue that neither of these is adequate to the problem of motivation: the former because it rests on a conceptual incoherence; the latter because its odds of practical success are vanishingly slim. I then turn to Locke, Rawls, and the idea of religious pluralism as a source of stability. First, though, some explanation is in order to flesh out the problem of motivation to which I have alluded.

4.1 Impartiality’s Motivational Deficit: The Problem of Ineffective Endorsement

It is worth saying something about why the issue of motivation is of particular concern to a political framework characterized by impartiality, in contrast to other kinds of political principles. The reason, in short, is that principles of impartiality are, by virtue of the rationalistic and abstract mode in which they engage citizens, limited in their ability to reliably motivate people. Liberal impartiality possesses inherently what has been called a “motivational deficit.”

1 Though she is not the only one to use it, I lift this phrase from Sharon Krause, “Desiring Justice: Motivation and Justification in Rawls and Habermas,” Contemporary Political Theory 4 (2005): 363–85.
The basic fact underwriting this deficit is the reality that compliance with a political order, while collectively rational, may, in some instances, be individually irrational. This is a basic coordination problem, and appears in all kinds of instances where citizens are reasonably tempted to cheat on their obligations: reporting their income for tax purposes, recycling, shoveling their sidewalk after a snowstorm. Everyone should rationally prefer a society in which everyone fulfills these responsibilities, but each individual can also improve her individual payoff by free-riding on everyone else’s compliance. We discussed this problem in Chapter 2, and suggested that it poses a challenge that modus vivendi stability is impotent to overcome. Instead, I argued, stability requires that citizens judge from within their conception of the good, that their balance of reasons tilts in favor of maintaining their support for their regime, even in the face of temptations to defect.

Paul Weithman dramatizes the free-rider problem in an updated rendition of Rousseau’s classic stag hunt dilemma. Rousseau described two hunters who collaborate in pursuing a stag, though either could break away and secure himself a hare. Each must choose between the greater, though uncertain, payoff of the stag and the lesser, though guaranteed, payoff of the hare.² Weithman’s scenario involves two machine gunners who hold dangerous advance positions in a battle.³ Each could save his own life if he flees, though it would mean sacrificing the life of his comrade and their objective in the battle. Rational calculation of each’s individual self-interest points toward running, despite the fact that each would prefer, all things considered, that the battle be won. Individually self-interested calculus, therefore, is inadequate to guarantee the collectively rational outcome. Applied to the political context this means that citizens will be

routinely tempted, when reasoning on the basis of their individual interest, to non-comply with their political obligations, thereby posing a threat to stability. Acting in support of a regime will always impose certain costs, and these will sometimes be severe. As Charles Taylor notes, any democracy will require “that its members be motivated to make the necessary contributions: of treasure (in taxes), sometimes blood (in war)…” Given the costs of citizenship, even persons who recognize the rationality of compliance may sometimes be swayed to do otherwise.

Gerald Gaus has described the problem as one of ineffective endorsement. Even cooperative schemes that actors have sufficient reason to endorse are not guaranteed to reliably and effectively motivate compliance on the part of those actors. The discussion in Chapter 3 was dedicated to identifying some necessary conditions for a regime to be capable of winning the endorsement of the citizens living under it. But we know that people do not experience the world through lenses of perfect rationality, nor are they motivated exclusively (or even predominantly) by rational calculation. This fact is borne out by evolutionary psychology, behavioral economics, as well as the everyday actions of gamblers, lovers, and soccer fans. People are motivated, rather, by their particular aims and attachments – the projects they care about, the desires that excite them, and the people they love.

This is where it becomes clear why the problem of ineffective endorsement is particularly salient for a political order governed by principles of impartiality. Such an order asks persons to ask on the basis of considerations and principles that are highly generalized, abstract, and rationalistic. The principles of impartiality call for political arrangements to disregard many of the aspects of individual existence that define a person’s distinctiveness – and which are likely to account for a great deal of what motivates her actions. Citizens will not be drawn to comply with

---

such principles by any particular, personal, and affective connections with them. The worry, in fact, is that they will not be drawn to comply with such principles at all. Principles are not the sort of thing that tend to tug at one’s heart or to stir one to action (especially in the face of countervailing considerations). Martha Nussbaum captures the point with characteristic succinctness: “People really don’t fall in love with abstract ideas as such, without a lot of other apparatus in the form of metaphor, symbol, rhythm, melody, concrete geographical features, and so on.”

An appeal to the abstract desirability of distributive and justificatory impartiality is likely to be insufficient to reach citizens in the ways that touch, inspire, and move them such that they would be motivated reliably to support the regime.

Nussbaum dramatizes impartiality’s inability to compel through the story of Marcus Aurelius, who dedicated himself to ridding his actions and judgments of all partiality. He learned to regard all persons uniformly, regardless of personal connections, and even forced himself to think of sexual love as “the rubbing together of membranes.” If this seems like a sadly denuded characterization of physical intimacy, that is exactly Nussbaum’s point: “What [Marcus] has seen is that impartiality, fully and consistently cultivated, requires the extirpation of the eroticism that makes human life the life we know.” Impartiality is synonymous with the sidelining of passion. Political regimes that are defined by impartiality thereby eschew some of the core phenomena that excite real human beings, and that drive real human behavior.

The problem of ineffective motivation is one that some observers have highlighted as a persistent, if not characteristic, issue within at least some corners of liberal political theory. Nancy Rosenblum makes the point, in broad terms, that the marginalization of affect is part and parcel of liberal impartiality. She writes that, “traditionally liberalism has warded off everything

---

7 Ibid., 225.
affective, personal, and expressive. That is the promise of impersonal government, and that is
what the discipline of tolerance and impartiality requires.”8 Martha Nussbaum has characterized
the problem as one of “watery motivation.”9 She draws the phrase from Aristotle’s criticism of
Plato’s ideal city, which attempted to eliminate partiality by erasing family ties. By asking
citizens to care equally for all other citizens, foregoing preference for their own kin, Plato’s city
is likely to generate weak – “watery” – motivation. This city fails to draw on what Aristotle
identifies as the most potent sources of effective allegiance toward some object: “the thought that
it is theirs and that it is the only one they have.”10 This is the basic problem of much liberal
political philosophy, Nussbaum claims. Theories that emphasize impartiality and adherence to
principles of universal morality (cosmopolitan views are a good example) undermine the
sentiments of particularistic attachment that can ground more than mere “watery motivation.”

One place to observe the challenge of motivating adherence to norms of strict impartiality
is in Habermas’s theory of discourse ethics. The motivational problem arises for Habermas as a
consequence of his overwhelmingly rationalistic account of moral reasoning. For Habermas, the
primary norm governing public justification is found in his Discourse Principle, which stipulates
the following: “Just those action norms are valid to which all possibly affected persons could
agree as participants in rational discourses.”11 Habermas’s aim is to define a deliberation
procedure that is entirely uncoupled from appeals to notions of the good. In order for a norm to
be valid, it has to be consistent with a purely procedural account of legitimacy, given by the
nature of rational discourse. Habermas specifies that a valid norm’s consequences and side

---

9 Nussbaum, Political Emotions, 219–225.
effects (given normal compliance with that norm) must be “such that all affected can accept them freely.” On his view, it is insufficient that a political claim be defensible on personal or private grounds; it must be capable of winning agreement from a wide diversity of perspectives. This implies, for Habermas, that some kinds of reasoning are not appropriate to the public square. For instance, he advocates a process of “translation” for adapting religious claims into a form that is publicly acceptable.\(^\text{13}\)

This emphasis on an abstracted, highly generalized conception of validity introduces a motivational deficit that Habermas himself recognizes.\(^\text{14}\) The reason for the deficit is that rational acceptance of a proposition or principle is, on its own, only a “weakly motivating force.” \(^\text{15}\) Under the abstract conditions of pure deliberation, he writes, “moral knowledge becomes detached from moral motivation.” \(^\text{16}\) In a discussion of this aspect of Habermas’s thought, Sharon Krause captures the issue in terms that point up the problem of motivation for liberal impartiality more generally: “Just norms may require us to forsake our more particular interests, our deepest ethical convictions, and key aspects of our personal identities, and when they do so the general interests served by justice may not be experienced as subjectively more compelling than these other factors.” \(^\text{17}\) Any hope of compliance with valid norms must rest not solely on their validity, but on some other factor drawing people to those norms and motivating them to resist temptations to defect from them. Consequently, Habermas is required to introduce into his view elements of affectivity and orientation toward the good in order to make his account of discourse valid.

\(^{15}\) Habermas, *Between Facts and Norms*, 147.
\(^{16}\) Habermas, *The Inclusion of the Other*, 34.
ethics motivationally plausible. These elements represent qualifications of one degree or another to the pure proceduralism of Habermas’s view. We will consider these in somewhat greater detail below. For now, the relevant thing is to note that the motivational limitations of moral/political principles that prescribe personal detachment, procedural rationality, and impartiality are recognized even by an arch defender of such principles.

A similar motivational deficit was influentially attributed to John Rawls’s theory of justice by Michael Sandel, in his early critique that inspired many communitarian critics of Rawlsian liberalism. Sandel objected to Rawls’s device of the original position on grounds that it supposed that real persons could – and ought to – reason politically from behind a "veil of ignorance" obscuring key facts about themselves and their commitments, and that persons could be reasonably expected to act in accordance with principles determined on the basis of such a depersonalized deliberation. While this critique took root in the imaginations of many of Rawls’s objectors, it rests on the basic mistake of assuming that parties to the original position are supposed to be real persons, rather than “artificial persons” best conceived as representatives or trustees of persons. Moreover, imputing this motivational deficit to Rawls’s work overlooks the great deal of attention paid to motivation in Part III of Theory of Justice. That said, the fact that Rawls’s theory does require some explanatory account of how real people will be motivated to comply with the principles selected by abstracted, impersonal contractors suggests that a theory of liberal impartiality does need to attend to the issue of motivation. (As it happens, I will argue below in favor of a basically Rawlsian approach to resolving this issue.)

---


The motivational limitations intrinsic to impartiality set up a dilemma. As we saw in Chapter 3, a regime must satisfy the requirements of distributive and justificatory impartiality in order for it to be reliably the case that citizens judge that their balance of reasons tilts in favor of maintaining their support for that regime. Impartiality is a necessary condition of stability. At the same time, impartiality suffers a motivational deficit. Political principles characterized by impartiality ask political actors to act on behalf of norms that are abstract, highly general, and rationalistic, but these are not the kinds of things that spur real people to action. If citizens are only weakly motivated to follow these principles, they will be easily swayed to non-comply, and stability will be undermined by the problem of free-riding. Impartiality is, therefore, both necessary for stability and potentially undermining to it.

What all of this means is that our account of liberal impartiality must attend to the problem of motivation. As we discussed in Chapter 2, basic facts of human psychology must be taken into account in the construction of a normative theory, and the fact that people are motivated most often and most forcefully by factors other than strictly rational calculation falls squarely in this category. A satisfactory theory of impartiality must not only account for how a regime can secure the reflective rational endorsement of those who live under it, but also must show how it can effectively motivate flesh-and-blood humans.

In his discussion of the “mortarmen’s dilemma,” Paul Weithman suggests that the soldiers might be motivated to hold their position if they were committed to a code of honor that forbid them to abandon the fight and their fellow soldier. This kind of devotion could counteract and, hopefully, outweigh their temptation to act on a narrowly self-interested calculus.20 Our question here is: what is the political analogue to a code of military honor? What force can

---

20 Weithman, Why Political Liberalism?, 52.
provide citizens with the necessary motivation to resist temptations to cheat on their political obligations and instead uphold principles that satisfy the requirements of impartiality?

4.2 Civic Identity

One place to look for a source of stability is in a shared identity that may unite citizens and motivate them to prioritize their political obligations over other of their interests or commitments. A classic example of this strategy is found in Rousseau’s account in *On the Social Contract*. I will argue that his proposals there are both unworkable and undesirable, but that a more plausible version is found in Jürgen Habermas’s idea of constitutional patriotism. But even Habermas’s account, I will argue, rests on an ideal that is basically untenable. First, though, we turn to Rousseau.

Rousseau’s account of the social contract offers strategies for grounding citizens’ commitment to impartiality in shared substantive commitments that are pre-political in nature, though which can be encouraged by the political order. Before identifying these strategies in detail, some exposition of impartiality in Rousseau’s contract theory is helpful. *On the Social Contract* is an analysis of the norm of perfect impartiality applied to political society. Rousseau’s starting point is what he calls the “fundamental problem,” which is to:

Find a form of association that defends and protects with all common forces the person and goods of each associate, and by means of which each one, while uniting with all, nevertheless obey only himself and remains as free as before.21

---

Rousseau’s interest is to reconcile humanity’s natural freedom with the authority that necessarily accompanies life in society. The solution he advances is that all members of society submit to a political order governed by the general will: “Each of us places his person and all his power in common under the supreme direction of the general will.” The general will is a component of Rousseau’s contractarian doctrine, and specifies the way in which an individual’s will may be identified with that of the collective, such that in obeying that will one may “obey only himself.” Rousseau does not supply a straightforward and unambiguous account of what the general will consists in, or how specifically it is to be determined, so readers are required to exert some exegetical effort in making sense of it. Gopal Sreenivasan has advanced the most analytically precise interpretation of Rousseau’s general will, which he interprets in largely procedural terms. In general terms, Sreenivasan characterizes the general will as “the constrained deliberative decision of the community,” by which he means that a collective deliberation process subject to certain conditions can be guaranteed to promote the critical interests that individual members of the collective hold in common. Sreenivasan identifies four necessary conditions that a deliberative process must meet in order to promote the common interest, and therefore to reflect the general will. They are:

1) the subject matter of the deliberation is perfectly general;
2) the conclusions of the deliberation apply equally to all the members of the community;

---

22 Ibid.
23 Gopal Sreenivasan, “What Is the General Will?,” *The Philosophical Review* 109 (2000): 554. David Lay Williams names Sreenivasan’s reading as an example of the most procedural possible interpretation of the general will. As we will see, though, this is not quite correct. Sreenivasan’s conditions for the embodiment of the general will encompass substantive requirements. For instance, he stipulates that the subject matter of the deliberation must be general. This is weakly substantive, but goes beyond mere constraints on procedure.
24 Ibid., 554–55.
3) all the members of the community participate in the deliberation;

4) all parties to the deliberation think for themselves.\textsuperscript{25}

The first of these specifies that proposals under consideration must not name particular individuals or their interests. They must be formulated in terms that are general, rather than singling out individuals or groups thereof.\textsuperscript{26} As Rousseau puts it, “There is no general will concerning a particular object.”\textsuperscript{27} Additionally, collective decisions must apply to all members of the collective: they may not be imposed on some but not others. There can be no exemptions from the scope of the decided-upon rule. These two constraints, taken together, require that a collective decision “must be general in its object as well as in its essence; that it must derive from all in order to be applied to all; and that it loses its natural rectitude when it tends toward any individual, determinate object.”\textsuperscript{28} Further, a decision consistent with the general will must be arrived at by procedures in which all participate. Rousseau describes the determination of the content of the general will in terms of a voting procedure that is equitable and open to all.\textsuperscript{29} But at the same time, Rousseau makes clear that the general will is not a purely procedural or formal notion: it consists in more than the preferences of the majority. He writes, “What makes the will

\begin{thebibliography}{9}
\bibitem{25} Ibid., 565.
\bibitem{26} Ibid., 566–67.
\bibitem{27} Rousseau, \textit{On the Social Contract}, 37. Richard Boyd writes that “Generality requires laws to be abstract and indifferent to the particular identities of the persons making them as well as those to whom they will apply.” “Justice, Beneficence, and Boundaries: Rousseau and the Paradox of Generality,” in \textit{The General Will: The Evolution of a Concept}, ed. James Farr and David Lay Williams (New York: Cambridge University Press, 2015), 250. This formulation captures the indifference of decisions of the general will to particular \textit{identities} (put another way, they must satisfy the anonymity condition well-known from social choice theory) but neglects that such decisions must also be indifferent to \textit{interests} that are not held in common. For instance, that my neighbor being prohibited from erecting a garish fence serves my interest is insufficient for its inclusion in the general will; it would need to be the case that \textit{all} members are served by a rule prohibiting garish fences in particular.
\bibitem{28} Rousseau, \textit{On the Social Contract}, 33.
\bibitem{29} Ibid., 82.
\end{thebibliography}
general is not so much the number of votes as the common interest that unites them.”30 The fourth constraint reflects Rousseau’s skepticism toward “partial societies,” or special interest coalitions that shape individuals’ reasoning and preferences.31 As Sreenivasan puts it, the operation of these coalitions “perverts the deliberation of the community” and can “serve to prevent the community's deliberation from promoting [the common interest].”32

Each of these constraints could be subjected to further scrutiny and elaborated in greater detail. For our purposes, though, this brief characterization of the idea of the general will is sufficient to demonstrate that Rousseau’s prescription for a society of perfect freedom is a regime of perfect impartiality. Persons can be free – obeying only themselves – by living under a rule that regards individuals and their non-shared interests with indifference.

So far, the concerns motivating Rousseau’s account of the general will appear to share more in common with the “moralistic” ends of equality and autonomy than with the Hobbesian priority of stability. This much is true; Rousseau is not concerned principally with stability. But neither is Rousseau blind to the problem of stability. A regime that satisfies the stringent norms of impartiality that the concept of the general will implies will face the motivational deficit that we identified above. Rousseau recognizes that the stability of the general will requires not only that citizens’ interests be served, but moreover that citizens identify with their society and genuinely desire the promotion of the common good. Sreenivasan casts this in the following terms: citizens must believe that collective decisions promote their critical interests.33 Unless citizens believe that they are served by the general will – independent of whether they in fact are – they will not identify with it. This failure of identification undermines the full generality of the

30 Ibid., 34.
31 Ibid., 32.
social will. Thus, the two conditions – one objective (that collective arrangements serve each citizens’ critical interests) and one subjective (that citizens believe that collective arrangements serve their critical interests) – are mutually reinforcing, or mutually undermining. Thus, some additional factor is required to motivate citizens’ identification with the community. Rousseau offers two possible such factors.

The first is an appeal to an ethno-nationalist identity that binds citizens to one another and grounds a sentiment of common interest. Persons who identify as part of a coherent national collective will be motivated by a sense of patriotism to maintain their commitment to the general will. This is a theme hinted at in On the Social Contract but more fully developed in other of Rousseau’s writings. In his Discourse on Political Economy and Considerations on the Government of Poland, Rousseau advocates the public cultivation of a sense of national distinctiveness and civic solidarity. These festivals, holidays, and rituals create from a collective of persons a people, set against others by the creation of a “national physiognomy.” The citizens of one state are to be discouraged from trusting persons other than co-nationals, and the distinction is to be so firmly drawn that assimilation is impossible.34 By directing a program of public education toward the cultivation of national identity, a citizen will develop loyalty toward “only his fatherland; he lives only for it; when he is alone, he is nothing.”35 Importantly, Rousseau calls for the active cultivation of these patriotic sentiments: he does not think that persons’ natural faculty of compassion is suitably reliable. Natural compassion is not bounded by political affiliations, so patriotic partiality must be created and encouraged.36 (This distinguishes Rousseau’s ethno-nationalist source of stability from accidental stability. Rousseau does not rely

35 Ibid., 189.
on the happenstance existence of some pre-political stabilizing factor, but rather describes a regime that generates the conditions for its own support.)

Some authors have tried to minimize this aspect of Rousseau’s writings. Joshua Cohen, for instance, argues that Rousseau’s ethno-nationalist appeals are logically superfluous to his account of the general will, and suggests that Rousseau may not even have sincerely endorsed his “most exaggerated statements about ethno-national solidarity.”37 Cohen recognizes, though, that “no sensible interpretation can put [these elements of Rousseau’s thought] to the side.”38 As Richard Boyd writes, Rousseau’s account of the general will presupposes that “political institutions must either presuppose an antecedent sense of civic solidarity (where present) or actively work to cultivate it in the hearts of citizens (where absent), sometimes by appealing to ‘communitarian’ narratives of civic solidarity,” or else risk citizens’ failing to identify with collective decisions regarding political arrangements.39

The second resource Rousseau identifies as a potential support to the stability of a society of the general will is a conception of civil religion. Rousseau affirmed the importance of religious toleration, so did not advocate total religious uniformity in society, but did suggest that a particular kind of shared religious identity could serve to reconcile the private identities of persons with their social identities as citizens.40 The conception of civil religion that Rousseau embraces does not encompass a complete comprehensive doctrine, but would rather establish some minimum shared beliefs for members of a given society. He has in mind a “purely civil profession of faith, the articles of which it belongs to the sovereign to establish, not exactly as

37 Cohen, Rousseau, 22.
38 Ibid., 5.
dogmas of religion, but as sentiments of sociability, without which it is impossible to be a good
citizen or a faithful subject."\textsuperscript{41} Rousseau’s conviction was that Christianity is incapable of
motivating the social solidarity that a political community requires,\textsuperscript{42} but that a form of civil
religion, designed with political aims in mind, could perform this necessary function.

Might these two resources – ethno-national solidarity and civil religion – provide the
motivational factor that Rousseau’s society of the general will requires for its stability? The
suggestion is that they can effectively motivate citizens to identify with the collective and to
embrace the general will as their own. Could they provide the conditions for citizens to feel, as
Aristotle suggested, that their political society “is all theirs,” and hold tight to “the thought that it
is the only one they have,” such that they would be reliably motivated to support it and comply
with its requirements? Could these Rousseauian strategies resolve the problem of liberal
impartiality’s motivational deficit? There are, unfortunately, some problems with these
approaches.

The first, and potentially most damning, issue is that Rousseau’s proposals require
finding a middle ground between appeals that are either too “thin” to motivate allegiance or too
“thick” to qualify as truly general, while it is conceptually unclear whether such a ground can
even exist. Let me unpack this claim. In order for it to be the case that a political order is
governed by the general will, the subject matter of its laws and institutions must take the form of
highly generalized rules – otherwise it is not a truly general will, and fails to solve the
“fundamental problem” of freedom. At the same time, Rousseau’s proposals of nationalist
solidarity and civil religion require that citizens identify themselves with some particular
identity, either national or religious. This sets up the balance. Lean too far in the direction of the

\textsuperscript{42} Ibid., 97–99.
general, and citizens will fail to identify with the regime; lean too far in the opposite direction, introducing particularistic subject matter into political identity, and the generality of the political order will be undermined.

Balancing acts may be difficult, but they are not always impossible. In the case of Rousseau’s strategies, however, there is reason to doubt even the possibility of achieving the proper balance between generality and particularity in political identity. Consider, first, the program of civil religion that Rousseau outlines. It is clear that Rousseau is attempting to walk a fine line in his characterization of the beliefs that will be inculcated as part of civil religion. While he asserts that this religion is a “purely civil profession” expressing merely the “sentiments of sociability,” the dogmas he stipulates include: “the existence of a powerful, intelligent, beneficent divinity...; the life to come; the happiness of the just; the punishment of the wicked; ...” These dogmas, despite Rousseau’s claims to a “purely civil” doctrine, are ineliminably theological commitments, no matter how minimal they may appear. As a consequence, some citizens will be unable to affirm the existence of a deity thusly specified, or may disagree with the eschatological statements regarding the afterlife. These persons will therefore remain outside the fold of those motivated by the tenets of Rousseauian civil religion. The doctrine he describes is too “thick” with content to succeed in winning the identification of all citizens. Of course, one could finesse Rousseau’s account to even further minimize the substantive content of this civil creed. For instance, it is likely that appeals to the afterlife could be excised without losing the motivational force entirely, and the theistic commitments may be alterable as well. But, the further one moves in this direction, the less compelling becomes the “religion” that remains. Indeed, a civil religion that is sufficiently general in its content so as to

---

43 Ibid., 102.
44 This assumes, of course, that citizens will disagree over religious matters – in other words, the fact of pluralism. Some comments on this assumption follow below.
become an object of identification for all citizens will lack the substantive objects of allegiance that are needed to effectively motivate that identification.

The same problem obtains for the appeal to ethno-national solidarity. If the regime cultivates a “national physiognomy” that is very robust, then it will be too exclusive to win the endorsement of all citizens, in which case it will become internally divisive and undermine the common identity of the populace. Some recent historical examples illustrate the issue well. Consider comments made in recent years by some public figures in the U.S. (such as Sarah Palin or Mike Huckabee) suggesting the existence of a “real America” – a notion meant to encompass rural and small-town demographics (i.e., conservative voters). The implication is that metropolitan centers, particularly on the coasts, are somehow inauthentically American. While it may be tempting to dismiss such talk as deliberately unsophisticated political rhetoric, it nevertheless reflects an underlying conception of “American-ness” that is robust enough to exclude significant sectors of the U.S. citizenry. Indeed, persistent questions about Barack Obama’s nationality and/or religion demonstrate that some conceptions of American-ness are restrictive enough to exclude even the President! This suggests not only that some conceptions of national identity are problematically exclusive, but moreover that thick conceptions of national identity will always be vulnerable to exploitation by elites looking to divide the citizenry against itself. A conception of national identity must therefore be inclusive enough to win the allegiance of all citizens and to preclude dividing the citizenry against itself.

Before turning to consider the merits of a more inclusive conception of national identity, note a couple additional facts that trouble the appeal to robust national solidarity as a source of motivation. First, a fact I alluded to in section 3.3: it is most likely practically impossible to artificially achieve uniformity of either ethno-national identification or religious belief. Given the
worldwide reach of communication media, the ready availability of international transportation, and the pervasiveness of global commerce, a proposal grounding stability in full social consensus around national identity and religious belief seems thoroughly naïve. If uniformity is indeed elusive, then Rousseau’s approach is destined to run aground, as citizens will affirm non-congruent ethno-national loyalties and religious beliefs, thereby undermining the motivational efficacy of the publicly-approved doctrines and, moreover, causing offense to those who dissent from the public orthodoxy.

Second, even if uniformity of religious and ethno-national identity were somehow to be achieved, it is not clear that this would represent a satisfying resolution to the problem that this dissertation has attempted to address. Specifically, we have here been concerned to explore the challenges presented to stability by the fact of pluralism. We have assumed the inevitability of disagreement, and are aiming to identify the ideal frameworks for achieving stability in the face of that disagreement. The Rousseauian strategies, in contrast, buy stability at the price of pluralism. Rather than resolving pluralism’s challenge to stability, civil religion and ethno-nationalist solidarity secure stability by stamping pluralism out. Whatever the independent merits of this approach may be, it represents an evasion of, rather than a solution to, the problems we are confronting here.

Joshua Cohen has advanced an interpretation of Rousseau that he thinks can leave behind Rousseau’s “communitarian” appeals to nationalism and patriotism. Cohen rejects the interpretation of Rousseau that would call for totally subsuming the private self into the public self, and instead sees room for significant disagreement and diversity in the Rousseauian society.45 The trick, on Cohen’s reading, is not to create an artificial conformity, but rather to train the human instinct of *amour-propre* to take egalitarian form. This can be done, he thinks,

by egalitarian practices: the experience of treatment as an equal will encourage a commitment to 
reciprocity, which will lead citizens to support impartial laws institutions.\textsuperscript{46} The civic unity
which will support the society of the general will, then, is not something to be layered on top of a
political order, but rather will arise as a product of citizens’ interactions under an egalitarian
political order. As Cohen writes,

“Extended to include all members of society, reciprocity generates a concern for the good
of all. That is, if each other member of a collection of persons displays a concern for my
good, then – assuming reciprocity – I develop a concern for their welfare in addition to a
concern for my own. … The experience of being treated as an equal in public arenas,
then, leads to the formation of a motivation that expresses this equality.”\textsuperscript{47}

This account of a commitment to reciprocity that arises out of life in a society of the
general will holds promise to ground motivation without requiring the anti-pluralistic moves of
Rousseau’s civil religion and ethno-nationalism. That said, there are reasons to doubt the success
of Cohen’s interpretation. For one, it is not clear that it stands as a fair reading of Rousseau.\textsuperscript{48}
More importantly, though, Cohen’s confidence in citizens’ organically progressive dedication to
reciprocity runs into the problem at the heart of impartiality’s motivational deficit: it requires that
citizens develop steadfast loyalties to principles, which rarely win citizens’ deep affection. That
is, even if citizens’ interest-oriented support of norms of reciprocity can be transformed into

\textsuperscript{46} Ibid., 124–27; 152–56.
\textsuperscript{47} Ibid., 154–55.
\textsuperscript{48} Dennis Rasmussen, for one, argues that Cohen is mistaken to think that amour-propre, as Rousseau understood it,
can be developed in an egalitarian, “non-inflamed” manner. See his “Review of Rousseau: A Free Community of
ethical commitment to reciprocity,⁴⁹ this commitment will in most cases remain vulnerable to the problem of ineffective endorsement. Cohen’s interpretation of Rousseau, then, does not so much resolve the problem of liberal impartiality’s motivational deficit as restate it.⁵⁰

We need, then, a conception of collective identity that is both admits of significant pluralism and is also substantive enough to win the affection of those who participate in it. Perhaps the most prominent recent attempt to provide such an account is Habermas’s conception of “constitutional patriotism,” which consists basically in allegiance to the principles and practices of constitutional democracy.⁵¹ Affection for these norms and values can, on Habermas’s view, take the functional place of nationalist and ethno-nationalist identities as mechanisms for binding citizens together and to their political order. The substance of Habermas’s political theory, as we considered above, rests on universal norms of legitimacy and generalizability, so something more particularistic is necessary to motivate citizens’ identification and allegiance. Constitutional patriotism offers a substantively thinner basis for national identity than Rousseau’s “national physiognomy,” grounding national identity in common commitment to the practice of democracy.

Perhaps unsurprisingly, constitutional patriotism has been criticized as being too thin to perform the function Habermas envisions for it. Rogers Smith, for one, sees in constitutional patriotism little improvement over standard liberal accounts of “civic nations,” in which identity

⁵⁰ Cohen does address the problems of free-riding and mutual assurance, but his answer has more to do with rule-making and institutional design than with routine compliance. He suggests that the way to avoid free-riding is to “structure institutions so that individuals and groups are not typically in a position to design and implement policies which can reasonably be expected to yield benefits for themselves alone and to impose the bulk of them on others” (pg. 160). This is well and good, but it does not address temptations to cheat once the laws have been established.
is based on “agreement upon certain political principles and procedures, supported by a universalistic constitute story of human personhood.”52 Along similar lines, Margaret Canovan describes constitutional patriotism as an “abstract” and “bloodless” conception of national identity.53 David Miller has expressed doubt that constitutional patriotism can functionally substitute for traditional nationalist patriotism. He notes that the principles around which constitutional patriotism is supposed to emerge “are likely to be general in form,” and that while subscribing to them might be constitutive of a certain kind of liberal political identity, “it does not provide the kind of political identity that nationality provides.”54 These critics converge in their skepticism that a procedural identity that consists in general principles of liberal democracy will be capable of motivating citizens’ allegiance. They would likely agree that “national identity has provided this common identity and trust, and that no other social identity in the modern world has been able to motivate ongoing sacrifices … beyond the level of kin groups and confessional groups,” such as are required by liberal political membership.55

Habermas’s response to this critique is that constitutional patriotism is not defined by allegiance to constitutional principles or values as such, but rather to constitutional principles and values as made manifest in some particular constitutional tradition. This conception of national identity does not require loyalty to “abstract” principles, but rather to the practice of liberal democracy in some specific time and place. Habermas writes that

“The political culture of a country crystallizes around its constitution. Each national culture develops a distinctive interpretation of those constitutional principles that are equally embodied in other republican constitutions – such as popular sovereignty and human rights – in light of its own national history. A ‘constitutional patriotism’ based on these interpretations can take the place originally occupied by nationalism.”\(^{56}\)

Constitutional patriotism is not, then, what Patchen Markell has called a “redirective” strategy, by which citizens’ affective attachments to their national identity is transmuted into attachment to universal principles.\(^{57}\) It is, rather, a commitment to a *particular* project of democratic citizenship and a *particular* interpretation of liberal constitutional principles.\(^{58}\)

This rebuttal, however, gives rise to another critique – this time from the opposite direction. If constitutional patriotism is directed not at universal principles but at their instantiation in particular institutions, then it cannot be separated from the traditional kinds of nationalism that Habermas wishes it to replace. One version of this claim holds that constitutional patriotism *assumes* a shared pre-political national identity. Bernard Yack, for example, alleges that “the audience for arguments about the focus of political loyalty is not some random association of individuals united only by allegiance to shared principles, but a prepolitical community with its own cultural ‘horizon’ of shared memories and historical experiences.” As such, Habermas's view “assumes the existence of the very prepolitical cultural community that he ... rejects in the name of a community based on rational consent and political

---

\(^{56}\) Habermas, “The European Nation-State,” 118.


principle.” 59 In order for constitutional patriotism to be possible, a democratic collective must already conceive of themselves as a national or cultural community.

A second version of the objection alleges that constitutional patriotism is indistinguishable from nationalism. The reason for this is that, as Habermas admits, constitutional patriotism must be directed at some particular constitution or constitutional tradition. What Habermas does not recognize, critics have suggested, is how this kind of affection will be inseparable from affection directed at the national and cultural identifiers of the collective. Any political society will have a majority culture, a majority language, and a set of dominant traditions and beliefs, and affection for that society’s constitution will inevitably run together with affection for those cultural elements. 60 As Maurizio Viroli captures it:

The object of the citizens’ love is their own particular republic; not just democratic institutions, but institutions that have been built in a particular historical context and are linked to a way of life – that is, a culture – of citizens of that particular republic. 61

Habermasian constitutional patriotism is therefore caught between two perils. Conceived in procedural terms, it requires allegiance to abstract principles, which are, as such, unlikely to stir the affections of real people. On the other hand, more substantive interpretations of national identity blur the distinction between constitutional patriotism and patriotism or nationalism of the conventional (non-general and exclusive) sort. In short, it is caught in the same bind as Rousseau’s doctrine of civil religion. One avenue of reply that Habermas has taken is to reject

these lines of critique as too heavily bound by the status quo. Simply because a fully civic patriotism has not appeared in history to this point, he suggests, does not mean that it could not be actualized. Habermas notes that European nation-states have arisen by varied historical processes, including some in which collective political identities arose prior to shared national identities – undermining claims that the former necessarily assumes or is dependent upon the latter. This suggests that the mechanisms for the development of constitutional patriotism do exist. Habermas writes, “If this form of collective identity was due to a highly abstractive leap from the local and dynastic to national and then to democratic consciousness, why shouldn't this learning process be able to continue?”

Habermas is surely right to dismiss any objections to his view that rely on an inductive fallacy: that because constitutional patriotism has not been realized to this point, that it will not or cannot be realized at some future time. Nevertheless, there remain reasons to doubt the viability of constitutional patriotism as a solution to the motivational deficit of liberal impartiality. One such reason is empirical: even if constitutional patriotism is conceptually possible, what are the odds of a national identity arising on terms that have not been observed in human history? If the claim of liberal nationalists is correct, and it is only nationalist identity that has been able to motivate the ongoing sacrifices that political membership requires, then it may be unreasonable to bank on the sudden emergence of an alternative, yet equally motivating collective identity. This is not a decisive objection, but it does cast doubt on the viability of this strategy for resolving impartiality’s motivational deficit.

There is an even more serious reason for doubt. Granting that motivationally efficacious affection for constitutional principles and practices is conceptually possible, the account of

---

62 Hayward, “Democracy’s Identity Problem,” 188–89.
constitutional patriotism cannot explain how citizens will be motivated to identify with and support their constitution and constitutional tradition, as opposed to some other. The values toward which constitutional patriotism is directed are universal principles that find specific interpretation in each (liberal democratic) country’s constitution. Absent a pre-political (i.e. cultural or nationalist) identification with a particular political society, it is not clear why a citizen’s affection for a constitutional tradition must be directed at her society’s constitution as opposed to another (liberal democratic) constitution. This is an objection that a number of authors have raised previously. David Miller, for instance, doubts that constitutional patriotism can “explain why the boundaries of the political community should fall here rather than there”⁶⁴ – in other words, why one ought to feel bound to this political community, rather than that one.⁶⁵

Anna Stilz has tried to answer this with a sophisticated account of democracy as collective action.⁶⁶ In short, she argues that democratic participation is required by virtue of a natural duty to create and/or support just institutions, which implies a duty to act as a democratic citizen. Fulfilling these duties requires that one act in concert with some group of particular others – i.e., fellow citizens in one’s political society. Stilz’s argument is interesting and worth consideration, but all we need note about it here is that it is something of a non-sequitur in relation to the objections we have just considered. Her argument does not show how citizens will come to feel bound to their particular political society and its constitutional tradition, but rather why they ought to. She notes in several places that her account is designed to demonstrate that democratic citizens have “a sufficient reason” for allegiance to their particular institutions.⁶⁷ But,

---

⁶⁴ Miller, *On Nationality*, 163.
⁶⁷ Ibid., e.g., 25, 113, 205.
as we established above, that one has a sufficient reason to comply with a set of institutions does not imply that one will be reliably motivated to do so. One’s recognition of the duties that Stilz articulates depends on their cognition of a rational argument and their being subsequently motivated by the force of that argument. As she writes, “a reflective understanding of our duties of justice is perfectly sufficient to ground our identification as citizens. … Nothing else – and in particular, no commonalities of language, culture, or history – is required.” The argument from democracy as collective action may answer the problem of particularity in regard to obligation, but not in regard to motivation.

One way of capturing this shortcoming of constitutional patriotism is by invoking the quotation from Aristotle that was employed above. While affection for principles of liberal democracy as instantiated in a particular constitution may lead an individual to feel that the political order under which they live “is all theirs,” it is unlikely to make them feel that “it is the only one they have.” As long as there are multiple liberal democratic constitutional orders in existence, they are equally candidates for the affection of constitutional patriots. A constitutional patriotism that does not rely on some pre-political source of national identification cannot ensure that citizens will be motivated to identify specially with the political order under which they live. There may well be successful arguments that supply citizens with sufficient reasons for allegiance to their own laws and institutions, but if rational arguments were adequate to motivate, then this entire discussion would be unnecessary.

All of this is to say that, despite various attempts to situate Habermasian constitutional patriotism as the bridge between universalism and particularity, the account remains caught between a too-thin attachment to universal liberal values that cannot reliably motivate and a too-thick nationalism that threatens division and exclusion. Even this subtle and sophisticated

---

68 Ibid., 205–6.
attempt to resolve the motivational deficit of liberal impartiality fails to articulate a conception of civic identity that is conceptually tenable as a strategy to mediate between the bloodless proceduralism of principle and the exclusive partiality of “communitarian” political allegiance.

4.3 Political Emotions

The Roussauian and Habermasian strategies we discussed strove to motivate loyalty (and, consequently, compliance) by specifying some shared identity that could generate affective attachment on the part of citizens. If citizens identify fully with the collective, then they are able to perceive the general will as identical with or at least a component of their individual wills, and will therefore be motivated to support political arrangements devised by a general will. Martha Nussbaum has advanced a noteworthy alternative approach to the motivational problem in liberal theory. In her diagnosis, a major problem for the liberal tradition has been a tendency to neglect the phenomenon of emotion. No society—indeed, no person—is untouched by the influence of emotion, but, Nussbaum charges, many articulations of liberal political philosophy—particularly Lockean and Kantian accounts—have not paid emotion attention commensurate to its ubiquity in social life.

This neglect is specifically problematic in light of the challenge of motivation. Liberals who have ignored emotion, she argues, have neglected the most potent motivational resource available to them. Her argument aims to correct this tradition of neglecting emotion by an account of how political societies can draw on citizens’ sentimental faculties to overcome the challenge of motivating allegiance and compliance. She strives to show how a political society can cultivate a “public culture of emotion” that reinforces individuals’ attachments to the political principles that govern their society, all within the framework of political liberalism. As

---

Nussbaum, Political Emotions, 3–6.
she describes it, she wants to answer the question: “How can a decent society do more for stability and motivation than Locke and Kant did, without becoming illiberal and dictatorial in the manner of Rousseau?”

This will involve, Nussbaum argues, encouraging both negative emotions – anger at violations of persons’ basic entitlements – and positive emotions – love and affection or the ideals embedded in the political system. Nussbaum’s approach, then, is less about finding a specific identity that citizens can share in common and that will motivate their allegiance, but about marshalling the resources of the political order to cultivate emotions that motivate citizens to support and comply with their political institutions.

One reason that some liberal thinkers have marginalized emotion is a fear that emotion functions in opposition to reason, and that affect poses a threat to rational action in the political realm. In particular, a certain tradition of Kantian thinking views emotion as a threat to a person’s ability to act impartially, and therefore a problem that must be overcome. Nancy Rosenblum has suggested that the prominent role of legalism in liberal thought reflects the belief that such an orientation “protects political society from the intrusion of emotional inclinations.”

The founders of the United States were well attuned to the “problem” of emotion, and sought a political order that would contain their ill effects. Madison, in Federalist #49, writes that “it is the reason of the public alone that ought to control and regulate the government. The passions ought to be controlled and regulated by the government.” Emotion, on this view, is a fact which must be acknowledged, but only for the sake of most effectively mitigating its social influence.

Nussbaum notes another reason for sidelining emotion in liberal thought, namely the concern

70 Ibid., 6.
71 Ibid., 6–7.
73 Rosenblum, Another Liberalism, 35.
that treating emotions under the rubric of the political is likely to lead in illiberal directions. She writes that that “liberal political philosophers sensed that prescribing any particular type of emotional cultivation might easily involve limits on free speech and other steps incompatible with liberal ideas of freedom and autonomy.”\textsuperscript{75} Liberals (again, principally those in the Kantian tradition) have feared that engaging with the emotional lives of citizens would require substantial interference in a sphere generally demarcated as private and off-limits to the legal-coercive mechanisms of the political order.

Not all liberals, however, have neglected emotion in the way that Nussbaum finds problematic. Indeed, she identifies within the liberal tradition some useful resources for her theory of political emotions. One is John Stuart Mill, who (influenced heavily by the thought of Auguste Comte) advocated a “religion of humanity” that could be taught in society as a substitute for existing religious doctrines.\textsuperscript{76} Mill’s idea was that the diversity of religious views in a given society pulled citizens in different directions, and that the stability of a liberal order could be best supported by replacing those views with a common system of belief supporting emotions conducive to personal sacrifice and general altruism. Of course, this strategy of religious homogenization would necessarily involve the inroads on individual autonomy that many other liberals have shied away from.\textsuperscript{77} For Nussbaum, a more attractive version of the religion of humanity strategy is found in the work of Indian intellectual Rabindranath Tagore. Tagore, like Comte and Mill, advocated a political program to encourage sympathy, and

\textsuperscript{75} Nussbaum, Political Emotions, 4.
\textsuperscript{77} It also relies, Nussbaum thinks, on untenable psychological and epistemological assumptions. Political Emotions, 74–81. Also, though I won’t explore it here, Nussbaum includes Wolfgang Amadeus Mozart among the antecedent models for her approach. Ibid., 27–43.
emphasized the use of the arts toward this end. In contrast with Mill, however, Tagore’s proposal is thoroughly pluralist, emphasizing individualism and constant questioning of tradition.78

Nussbaum also identifies John Rawls as a resource for a political theory of emotions. She writes that Rawls’s “sophisticated conception of emotions” in *A Theory of Justice* was “compelling and insightful,” and “ahead of its time.”79 The account that Rawls gave in *Theory* focused on emotional development within the family. He noted that children tend naturally to develop feelings of love and care for their parents. Rawls’s contention is that these emotions can be extended over time to the political principles embedded in political arrangements. As persons develop social ties within a just society, he suggests, their caring sentiments will come to include those other people. Eventually, as people develop greater appreciation for the just institutions under which they live, they will come to love those institutions and their underlying principles as well.80

Nussbaum is simultaneously critical and appreciative of this argument of Rawls’s. She ultimately does not find it persuasive, as it omits any compelling account of how this extension of love and care is likely to occur. Loving one’s parents is quite different from loving principles of justice, and Rawls’s confidence in the possibility of this leap is not well-explained. “He says nothing,” Nussbaum notes, “about how the particularity of these loves will actually lead on to a grasp of general principles.”81 Rawls himself seems to drop this account of extended sentiment in his later writings, but Nussbaum commends him for recognizing that a theory justice requires a

---

79 Ibid., 9.
81 Nussbaum, *Political Emotions*, 221.
“reasonable moral psychology” that can explain citizens’ motivation to support their regime.\(^{82}\) Nussbaum’s aim is to fill in this moral psychology by means of an account of political emotions.

The key to a politically successful treatment of emotion is that it must explain how a regime can counteract worrisome psychological tendencies that threaten the maintenance of a just society. The most pernicious such tendencies that Nussbaum names include, first and foremost, narrow sympathy, but also deeper issues such as disgust and shame at the human body, and tendencies toward submissiveness to authority and peer pressure.\(^ {83}\) A political society must proactively encourage citizens in developing psychological resistance to these pathologies, and must guide its citizens toward sentiments of love, instead. Nussbaum’s conception of love, in this context, is multifaceted; it includes:

“a delighted recognition of the other as valuable, special, and fascinating; a drive to understand the point of view of the other; fun and reciprocal play; exchange, and what Winnicott calls ‘subtle interplay’; gratitude for affectionate treatment, and guilt at one’s own aggressive wishes or actions; and, finally and centrally, trust and a suspension of anxious demands for control.”\(^ {84}\)

This love ought to extend, Nussbaum insists, to love for the nation and its history. This is the key to securing citizen’s stable commitment to maintaining the institutions that shape their political society. She tells her own story of, as a child, being captivated by the story of the American founding. She reveled in the stories of bravery and innovation in the formation of a political society. She tells her own story of, as a child, being captivated by the story of the American founding. She reveled in the stories of bravery and innovation in the formation of a

\(^{82}\) Rawls, *Political Liberalism*, 82.

\(^{83}\) Nussbaum, *Political Emotions*, 379.

\(^{84}\) Ibid., 176.
new nation, and it was through her affection for these stories that “the abstract values of liberty and individualism were eroticized” – that is, made into objects of her love.85

Among the mechanisms Nussbaum suggests for the cultivation of these sentiments are education (for teaching patriotism), as well as festivals, celebrations, and works of art to stir emotion. She prescribes a combination of tragic and comic festivals. In the former category fall the Gettysburg Address and Whitman’s poem commemorating the same battle, public photography projects funded by the New Deal, and the Vietnam War memorial in Washington DC. Comic festivals include Millennium Park in Chicago (with its diversity-celebrating art installations), and even the comedy of Lenny Bruce, who prodded American society to a “difficult conversation.” These public engagements spark emotional development in citizens, encouraging them along in hopes that “a love-infused compassion can become, extended, a vehicle of political principles.”86 The trick to solving the motivational problem in liberal impartiality is to tug at citizens’ heartstrings in public celebrations of art and music.

In keeping with her other work in political philosophy – most notably her formulation of the capabilities approach87 – Nussbaum strives to frame her political program of emotional education as fully consistent with the commitments of political liberalism. Namely, she maintains that the measures she advocates and the sentiments she wishes to cultivate are part of a conception of justice that can potentially become the object of an overlapping consensus among all reasonable citizens. This requires, Nussbaum recognizes, that political principles be both “narrow in scope” and “shallow as to their basis.”88 Correspondingly, the public emotions stoked

85 Ibid., 250.
86 Ibid., 313.
88 Nussbaum, Political Emotions, 129.
by art, festival, and education ought to be both narrow and shallow in comparison to the comprehensive doctrines that citizens hold individually and in association. We might compare here Corey Brettschneider’s recent suggestion that the state can “speak” affirmatively in favor of basic democratic values without betraying citizens’ right to hold anti-liberal comprehensive beliefs. Nussbaum calls for the state to do more than speak, though. She calls for the state to elicit from its citizens sentiments of love and affection directed at the political order sufficient to render it stable.

There can be no doubt that affectionate emotion toward one’s political community can aid in maintaining that community as stable over time. Nussbaum’s personal story of passion for narratives of the American founding rings true, and one can without difficulty observe that many others feel a similar brand of love, rooted in history and idealism, for the United States. To be sure, a great many people the world over feel love for their political societies. The truth in Nussbaum’s identification of emotion as a source of stability can be observed by noting the intuitive truth of its converse: when people become apathetic (or, worse, antipathetic) toward the political institutions under which they live, the stability of those institutions is weakened. And yet, there are limitations to the usefulness of political emotions as a support for stability.

For one, it’s not clear that the kind of stability supported by emotions of love would qualify as stability for the right reasons. Very often, love that a person feels for his or her political society is entirely accidental, in the sense we discussed in chapter 2. It may come about by virtue of the fact that a person’s family, friends, hometown, familiar topography, personal narrative, and/or ethnic group happen to be situated within the bounds of a given country. The love a person feels for his or her country may be entirely derivative of the love she feels for one

---

of these other objects. That her affection happens to extend to the political order is purely a matter of contingency and accident.

Of course, Nussbaum calls for doing more than simply taking advantage of supportive affection that happens to obtain; she advocates actively encouraging emotional attachment. But it is not clear why support achieved on the basis of deliberately cultivated emotion should qualify as a “right reason” for stability. For, let us call a spade what it is: Nussbaum is talking about emotional manipulation. She calls for the political order to strategically exploit its citizens’ emotional malleability. While this may be a practically effective means of achieving stability, it has more in common with stability achieved by (material) coercion than stability for the right reasons. Stability on the basis of emotional attachment may have nothing to do with the normative character of the regime; it is a product of a special emotion-targeting strategy by the state. Emotionally-engineered support is, in a certain sense, even more nefarious than support secured by brute coercion. When support is coerced, citizens at least have the opportunity to choose autonomously whether to comply or to bear the punishment. Emotional manipulation, in contrast, gets into citizens’ heads, and aims to preempt their rational calculations regarding whether to maintain their support or to defect. If strategic cultivation of citizens’ emotions can secure stability, it is not stability for the right reasons.

The more serious problem with Nussbaum’s approach is that it must strike the balance between appealing to content that is too thin, on one hand, and too thick, on the other, just as the civic identity proposals did. Nussbaum wants to find a middle ground between an implausible “love” for highly rationalistic principles of justice and crudely particularistic patriotism that is exclusionary and uncritical.\(^90\) Veering too far in the direction of one or the other amounts too failure. Affection that is too particularistic in its object runs the risk of national arrogance and

---

xenophobia, as well as the possibility that citizens’ affection will be directed at the wrong objects. Rather than taking the basic principles of the regime as the object of their love, they may instead become to devoted to their immediate social group, to people sharing in their “way of life” (however they conceive it), to their ethnic or cultural group, or even to some simplistic ideal of the nation’s history. Any one of these misdirected loyalties is potentially destabilizing, insofar as it would conflict with general political interests. On the other hand, if emotions are developed in a manner that leans too far in the direction of the abstract, they run the risk of either “watery emotion” or unmoored cosmopolitanism, wherein persons direct their sympathies not toward their own society and their compatriots, but rather toward humanity in general.

Nussbaum’s own description of how her program of emotional development is supposed to work demonstrates the inherent difficulty in eliciting the proper affective responses from a citizenry. She says that her account “aims to strike a delicate balance between too thick a theory and no theory, something that is not easy to do, but which must be attempted.” 91 This involves the dual task of encouraging affection and loyalty toward the political society while also developing citizens’ critical faculties and willingness to dissent, particularly concerning the ideal of inclusiveness. She wants children as they grow to “come to love an America, or an India, that really stands for inclusiveness.” 92 At the same time, she advocates teaching “historical truth” and love of “the nation as it really is.” 93 The former sounds suspiciously like engendering affection not for America or India, but rather for an ideal of inclusiveness; the latter sounds like patriotic loyalty in spite of violations of ideals like inclusiveness. The former describes a love not really directed at the political society but at a principle, while the latter describes a love that too easily loses its critical edge.

91 Ibid., 201.
92 Ibid., 252.
93 Ibid., 254.
Even the anecdote of Nussbaum’s childhood enthusiasm for tales of the American founding points up the problem. She claims that she came to love the American founding and identified herself with that drama, and that the abstract values of political liberty and individualism “acquired motivational force” through their embodiment in the characters she admired and her feelings of admiration themselves. She says that these abstract values “were eroticized” in this process. What she means is that she came to associate the principles with memories and experiences that she could latch onto affectively. She does not love the principles themselves, as such, but rather the constellation of concrete phenomena that have are associated with the principles. Indeed, it is only these phenomena – historical events, geographical sites, personal relationships – and not abstract principles that possess the motivational force Nussbaum is describing.

A program of political emotions must therefore direct citizens’ affections toward these concrete objects, working and hoping that the citizens associate them with the appropriate principles. Happily, Nussbaum’s affection for the founding has led her to affirm a critical, humanitarian patriotism. But an alternative history, in which her youthful patriotism evolved into a worrisome xenophobic nationalism, is equally plausible. The fetishization of the American founding by politically conservative Americans, exemplified by the Tea Party movement, represents the realization of this alternative. Since the principles of liberty and individualism themselves are not properly the object of patriotic affection, there is no guarantee that patriotic

---

94 Ibid., 250.
95 In fact, one is tempted to wonder whether Nussbaum’s cosmopolitan sensibilities don’t have significantly to do with concrete affections directed at people and places outside the U.S. and the American political tradition, as described in her Women and Human Development. That said, amateur psychoanalysis of an author with whom one has no personal relationship is vaguely presumptuous at best, and outright offensive at worst, so I will leave the suggestion at that.
emotion will take the critical and liberal form that Nussbaum hopes. It may instead develop into the narrow and exclusionary form of nationalism that we worried about above.

The problem with the civic identity proposals was that they relied on an elusive conception of shared identity that struck the proper balance between civic proceduralism and conventional nationalism. Both Rousseau and Habermas aimed at conceptions of identity that were at once fully general and also particularistic enough to secure allegiance. This amounts, though, to squaring a circle, and I argued that neither thinker’s proposals succeeded in doing so. Nussbaum, in contrast, does not invoke any specific conception of identity, so she may be able to sidestep this conceptual problem. But the same issue will plague her program in practical terms. The likelihood that a political order will be able to strike the extremely narrow and tenuous balance between too-thick and too-thin is small. The program that Nussbaum proposes requires that citizens not only develop affections for the prescribed objects, but that they form affections-by-proxy for the appropriate principles. Strong cognitive and affective temptations tug away from this mean in both directions. Evidence overwhelmingly points to the tendency of politically societies to miss the mark. What’s more, failure to strike the proper balance is likely to compound: as citizens grow attached to their particular place/people/nation/etc., abstract principles are likely to appeal to them less, and vice versa. The domain of success, in other words, is so narrow that it is practically unlikely to be achieved. Nussbaum wants to achieve a practical approximation of what Rousseau and Habermas are unable to identify in conceptual terms, but there is reason to doubt that her program of political emotions represents a plausible approach to navigating between the perils of “watery emotion” and exclusive nationalism.
4.4 Harmonization of Loyalties

Having cast some appreciable doubt on these two strategies for motivating allegiance to impartialist political regimes, we now turn to the approach that I think bears the most potential for resolving the motivational deficit of liberal impartiality. By no means have we considered all possible candidates, but the approaches taken by Rousseau, Habermas, and Nussbaum are worthy contenders, and they provide useful points of comparison in considering the alternative I endorse. The primary point of difference between my preferred approach and the previous two is that, where appeals to civic identity and political emotions relied on some form of cultivated homogeneity – either of religious/national identity or emotional dedication – to motivate support, the approach I prefer embraces pluralism and draws on the motivational force of citizens’ diverse moral and religious commitments to drive their support or the regime. The principal source I will draw on to illustrate this strategy is John Locke’s “Letter Concerning Toleration.”

Locke’s letter advocates impartiality both distributive and justificatory, but is also attuned to the problem of motivation. The text is aimed specifically at persuading persons who are firmly committed to their distinctive conceptions of the good (i.e. religious beliefs) that those commitments give them strong reasons to affirm an impartialist regime. It strives, in short, to “harmonize” citizens’ religious loyalties with their political ones.

My argument is that this Lockean approach to motivation is both superior to the other two approaches we considered as well as independently defensible. By this I do not mean that it is a panacea, a foolproof strategy for overcoming the motivational deficit of liberal impartiality. Indeed, I will make clear that it is decidedly not guaranteed to work always and everywhere; some social and political circumstances will render it practically useless. The harmonization of

---

loyalties approach shares this potential for failure with the civic identity and political emotions approaches. But, in this section and the next, I will argue that the harmonization approach stands above those other approaches in three important ways. First, it does not contain any fundamental conceptual incoherence, as did the civic identity approach. Second, the circumstances in which the harmonization strategy could work can be plausibly imagined to obtain in some actual political contexts. Third, the harmonization approach is amenable to imperfect realization. Failure of this approach is a matter of degree, and small deviations from perfect realization imply only small detriments to stability. I will explain each of these in greater detail in the discussion that follows. For now, this brief overview should suffice to give some sense of the standards by which I judge the harmonization strategy to be superior to the alternatives we have considered.

The key to recognizing the motivational dimension of Locke’s “Letter” lies in attending to the theological, and not merely political, aspects of the text. The political substance of the text, of course, is clear. He calls for a doctrine of religious toleration, which most straightforwardly entails a political policy of religious neutrality in which persons are treated impartially with regard to their religious beliefs.\(^{97}\) Locke’s argument relies on a distinction between the concerns proper to the magistrate, which are temporal and which can be enacted by force, and concerns proper to the church, which are spiritual and are not amenable to coercive enforcement.\(^{98}\) The state, Locke holds, is not capacitated to intervene in matters pertaining to religion, and therefore must not. This entails a program of strict impartiality with regard to the religious beliefs of citizens, from which Locke even derives a blanket denial of religious exemption. He writes, “For the private judgement of any person concerning a law enacted in political matters, for the public

---

\(^{97}\) Though not total neutrality: Locke is clear that atheists do not merit toleration. Ibid., 426. Locke’s argument is very interesting, but I will not devote space to it here. See chapter 10 of Jeremy Waldron, God, Locke, and Equality: Christian Foundations of John Locke’s Political Thought (Cambridge: Cambridge University Press, 2002).

good, does not take away the obligation of that law, nor deserve a dispensation.”

When religious beliefs motivate actions that contradict the laws of the state, religious persons are to be shown neither prejudice nor privilege. An action that is permitted when undertaken for secular purposes is permitted when undertaken for religious purposes, and an action prohibited for secular purposes is prohibited for religious purposes. The point is that when religious actors tread into the domain of civil authority, they become subject to that authority, just as a resident of Virginia who crosses into Maryland becomes subject to the laws of the latter. This jurisdictional account of religious toleration, relying on a distinction between proper and improper objects of state regulation, is the most prominent line of argument in Locke’s “Letter.”

The argument entails not merely distributive impartiality – treating citizens without respect to their religious beliefs – but also a degree of justificatory impartiality. Locke suggests that overtly religious rationales for coercive action are not reasons that persons subject to that coercion could rationally accept. He writes that toleration is necessary because “no man can so far abandon the care of his own salvation as blindly to leave it to the choice of any other...” As David Wootton summarizes, “It is not rational for subjects to hand over to their rulers responsibility for deciding what they should believe.” To employ the vocabulary of contemporary liberal theory, agents of the state who use coercive force to enforce religious precepts fail to act on reasons that all can accept. This conclusion is entailed by Locke’s religious epistemology, which seats the definitive religious faculty in the individual, rejecting the notion

---


that some persons are of sufficient religious expertise that they ought to direct the belief of others. 103 At the very least, Locke does not see any reason to privilege the religious judgment of the magistrate: “The one narrow way which leads to heaven is not better known to the magistrate than to private persons, and therefore I cannot safely take him for my guide, who may probably be as ignorant of the way as myself, and who certainly is less concerned for my salvation than I myself am.” 104

Religious belief, on Locke’s view, is a private interest, insofar as it is something to be adjudicated at the level of the individual, and over which no other person may be committed to compel others. Locke’s argument is intended most directly as a negative denouncement of established religion, but it implies a positive condition that state coercion should be accompanied by justifying reasons that do not depend on the truth of substantively religious premises. 105 This commitment is reflected in Locke’s approach to religious exemptions, in the sense that he takes religious beliefs as insufficient to justify either the laws themselves or exemptions from them.

So far, Locke’s argument for a political norm of impartiality would appear to have more to do with conditions of respect for individual conscience than with stability. Or, as some commentators have noted, Locke endorses toleration in large part as a condition for the possibility of individuals accepting Christian belief. 106 On either reading, Locke affirms toleration as a means to the protection of free conscience and belief. At least overtly, he is not concerned

103 “[I]t appears not that God has ever given any such authority to one man over another as to compel anyone to his religion.” Locke, “Letter Concerning Toleration,” 394.
104 Ibid., 408.
with political stability. But Locke’s argument for impartiality with respect to religion, while in substance it proceeds from apparently “moralistic” premises, reflects a rhetorical strategy that promises to help resolve the motivational deficit in liberal impartiality. As mentioned above, the key lies in recognizing the theological appeal embedded in Locke’s “Letter.”

The text opens by first declaring that toleration is the “chief characteristical mark of the true Church.” That is, Locke’s first move is not to show that religious toleration is a political priority, but rather to show that it is a theological priority. John Perry has argued that Locke recognizes the fact that individuals are subject to the demands of various loyalties that often compete with one another. Religious and political affiliations are chief among these, and the conflict between them can create the conditions of what Michael McConnell has called “citizenship ambiguity.” Locke’s jurisdictional account is designed to address the fact that, in practice, religious and political authorities do sometimes overlap, and this creates a difficult conflict for religious citizens. Perry calls this rhetorical strategy an effort in the “harmonization of loyalties,” in the sense that it tries to bring the requirements of one’s loyalty to the political order in line with the requirements of one’s religious commitments. Locke’s refusal of toleration to Catholics bears this out: their fidelity to the state cannot be trusted, for reason of their loyalty to religious authorities. This kind of suspicion toward Catholics persisted for centuries, though it appears to have been put to rest in the post-Kennedy era.

In the present day, examples of this conflict of identities can be readily observed in cases such as that of Kim Davis, a county clerk in Kentucky, who claimed that being required to issue marriage licenses to same-sex couples (following the Supreme Court’s decision in *Obergefell v. Hodges*) posed a conflict. Davis claimed allegiance to her constituents, but also to her religious commitments: “In addition to my desire to serve the people of Rowan County, I owe my life to Jesus Christ who loves me and gave His life for me.” To issue licenses to same-sex couples would force her to “violate a central teaching of Scripture and of Jesus Himself regarding marriage.” Kim Davis’s case made headlines because it involved a public official and a controversial social issue, but it merely dramatizes the latent conflict within many people of both religious and civic commitment. Locke recognizes the potential for this conflict, and thereby implicitly acknowledges the motivational problem: a straightforwardly political account of impartiality will not suffice to motivate support.

The harmonizing strategy rests on giving religious persons reason to believe that political impartiality with respect to religious belief is compatible with their own particular religious commitments. If this can be established, then enduring a political policy of toleration will not strike religious believers as a betrayal of their religious identity in deference to their political one, but rather will reflect the fulfillment of two harmonized loyalties. Under these conditions, the motivational deficit should be overcome, since citizens will draw motivation for support of an impartialist regime from their comprehensive doctrines. Locke’s argument is largely, as David Wootton notes, “about what Christians should believe.” His goal is to show that support for political toleration is in harmony with Christian belief, and therefore Christian citizens can – indeed, should – support norms of toleration.

---


245
Note a certain similarity of this approach to that advanced by Rousseau. Both resolutions to impartiality’s motivational deficit rest on an appeal to the individual beliefs – what Hobbes called the “private judgment” – of citizens. For Rousseau, it was only by seeing to it that citizens came to believe in the tenets of the civil religion that a society of the general will could sustain itself. Locke also looks to citizens’ religious beliefs to motivate support, but in a crucial departure from Rousseau’s strategy, Locke neither assumes nor insists on uniformity in citizens’ private beliefs. Rousseau proposed generating support by promoting a comprehensive doctrine that citizens could hold in common, but Locke takes a tack we could describe as decentralized. He endeavors to show that the comprehensive doctrines that citizens already hold do, in fact, support norms of impartiality. Locke (unlike Hobbes) does not attempt to revise Christian belief in order to render them supportive of the regime, nor does he prescribe a uniform code of belief supportive of the regime. Rather, by his justificatory strategy he implies a hope that a regime characterized by impartiality can win support from the perspective of the religious views held by the citizens it governs.

Locke wrote for Christians (a certain brand of Protestant Christians, more specifically), but we can extrapolate from the approach of his Letter to derive a strategy for justifying liberal impartiality in contexts of pluralism. The conclusion to be drawn is: impartiality can win effective support when it can win endorsement from the various religious beliefs populating a political society. If, for each religious viewpoint, toleration is a “characteristical mark” of that view, then a regime of toleration will have support from the citizens who live under it. If you will recall the discussion in chapter 2, you will note the similarities between this scaled-up Lockean move and the convergence interpretation of Rawls’s idea of the overlapping consensus. Rawls grounds the stability of a political regime in the endorsement of that regime from the
perspectives of citizens’ respective comprehensive doctrines. Rawls’s formulation of the overlapping consensus involved an additional level of agreement, however, in that the object of the consensus – a political conception of justice – was required to be articulable in a manner “freestanding” of any particular comprehensive doctrine. In section 2.4, I argued against Rawls’s belief in the necessity of a freestanding political conception – a conclusion also defended at greater length by authors such as Gerald Gaus and Kevin Vallier. Robust consensus on the values that justify a set of political principles is unnecessary; a mere convergence around those principles is sufficient for a regime to be stable (for the right reasons). This is what happens when each citizen comes to regard their religious and political loyalties as adequately harmonized.

We can now see the major ways in which Locke’s approach in the “Letter Concerning Toleration” differs from the strategies above. Rousseau and Habermas advanced specific conceptions of civic identity that they hoped citizens could rally around. Nussbaum proposed mechanisms for eliciting certain sentiments that would motivate citizens to support their political order and its underlying values. The Lockean strategy, in contrast, does not make any appeal to a particular identity or a particular set of sentiments that it hopes citizens will embrace. It leaves these matters at the level of the individual. As it happens, this indifference is something for which Nussaum criticizes Locke specifically. She writes that Locke

“made no attempt … to delve into the psychological origins of intolerance. He thus gave little guidance about the nature of the bad attitudes and how they might be combated. Nor
did he recommend any official public steps to shape psychological attitudes. The cultivation of good attitudes is left to individuals and to churches.”

The harmonization-of-loyalties strategy sees Locke’s avoidance of the psychological basis of tolerance as a strength, though, rather than a weakness. There is no need for the state to be involved in cultivating allegiance to the political order if it can draw on the sentiments that stem from citizens’ respective comprehensive doctrines. Individuals and churches powerfully shape the emotional allegiances that citizens feel, and Locke’s strategy draws on these as a resource for stability.

The major difference from the civic identity approach is that the Lockean harmonization strategy does not regard pluralism as a problem for stability. Rousseau judged diversity of beliefs to be a threat to the stability of a society of the general will, so he proposed programs of civil religion and ethno-nationalist solidarity in order to promote uniformity. Locke’s harmonizing strategy, in contrast, implicitly suggests that pluralism could be a strength: when the motivation for supporting impartiality comes from citizens’ respective comprehensive doctrines, it will effectively sustain citizens’ commitments to support their regime. It will be sufficient to overcome the problem of ineffective endorsement.

In sum, the strategy of drawing support from citizens’ commitment to their comprehensive doctrines provides a tenable strategy for overcoming impartiality’s motivational deficit. Whereas the strategies adopted by Rousseau, Habermas, and Nussbaum work at imprinting the values of political order onto citizens’ individual commitments (whether religious, patriotic, or affective), Locke’s strategy of harmonization draws on citizens’ deep values as they are to elicit support for the regime.

---

113 Nussbaum, Political Emotions, 4.
4.5 The Promise of Pluralism

The harmonization of loyalties strategy has the virtue of conceptual tenability: it describes an outcome that is possible, at least in theory. There is no a priori reason to deny the possibility of citizens’ comprehensive doctrines providing them with effective motivation to support and comply with their political regime. In this respect, it is preferable to the civic identity approach, which faltered in specifying a conceptually coherent version of national identity to which citizens could develop allegiance. The harmonization strategy is not incoherent in this way. The question remains, though, whether the harmonization of loyalties is a plausible approach to stability. This was the weakness we identified with Nussbaum’s program of political emotions: while there is no conceptual problem with her proposals, it is highly unlikely that they could be implemented successfully in any practical setting.

The plausibility objection is an obvious one to raise against the harmonization strategy. Even if some particular interpretation of Protestant Christianity can be shown to support religious toleration, as Locke intended, from this it clearly does not follow that all religious beliefs are supportive of, or even compatible with, toleration. Particularly in an age of fanatical religious violence and persistent conflicts in Europe (among other places) over religious versus political identity, it may seem beyond naïve to stand behind a premise affirming broadly the compatibility of religious belief and liberal impartiality. Not all religious beliefs are reasonable in the Rawlsian sense, and the assumption of pluralism must be understood to include these beliefs as well. In Nancy Rosenblum’s terms, some religious beliefs that will flourish in a liberal society will be “incongruent” with liberal values. Some citizens will hold religious views of a

---

theocratic bent, and will be unable, on the basis of those beliefs, to embrace political impartiality with respect to religious beliefs. So Locke’s harmonization of loyalties strategy would seem to be of limited applicability, as no such harmonization is possible for persons holding “unreasonable” religious views.

This objection rests on premises that are factually true, though it is directed at a distorted version of my claim. I have not intended to make the claim that all religious beliefs currently existing in the world do in fact support of liberal impartiality. This would be obviously false. Rather, the premise of the harmonization strategy is that a scenario in which citizens’ religious beliefs within a given society all come to support liberal impartiality is possible, and could be realized in reasonably plausible conditions. Even this moderated claim, however, requires some significant defense.

One assumption that the harmonization strategy requires is the following: that no religious tradition is, in principle, incapable of achieving the harmonization of loyalties – that is, of endorsing liberal impartiality. If this were not the case (i.e., if one or more religions were fundamentally and immutably opposed to either the ideals or the institutions of impartiality), then the harmonization strategy would be impossible to pull off. But there is ample reason to believe that religious opposition to impartiality is a matter of historical contingency, rather than a necessary element of religious views. Even bodies of doctrine that appear entirely opposed to liberal impartiality can, given the proper context and interpretations, come to embrace norms of toleration and neutrality. This is important because it grounds caution in writing off any particular religious group as inherently problematic vis-à-vis the stability of an impartialist regime.
A potent historical example can be found in the evolution of the Catholic Church in its stance regarding religious liberty and toleration. As recently as 19th century, the Church was loudly opposed to religious liberty, with Pope Gregory XVI, in an 1844 encyclical letter titled *Inter Praecipuas*, complaining that defenders of religious toleration “are determined to give everyone the gift of liberty of conscience, or rather of error.”115 Just over a century later, the second Vatican council produced a document (*Dignitatis Humanae*) devoted entirely to affirming religious freedom. This shift was part of a general Catholic warming to liberal values and democratic institutions, which Charles Curran describes in stark terms:

“The Catholic Church in the nineteenth century strongly opposed all forms of liberalism and regarded democracy as political liberalism. By the end of the twentieth century, however, the Vatican had become a world leader in advocating the cause of democracy throughout the world.”116

Skepticism about the compatibility of Catholic belief with loyalty to American political values and institutions persisted through the middle of the twentieth century – as evidenced by the need for John F. Kennedy to convince Protestants of his independence from the Pope in

---

advance of his election as president – but today is a mere artifact, difficult to take seriously. This transformation is evidence that even those religious doctrines that appear firmly and permanently opposed to liberal values such as impartiality are capable of transformations that make the harmonization of loyalties possible.

One condition for the occurrence of these kinds of transformations is a liberal political order where citizens of each and all religious groups are treated fairly and well, and therefore have reason to endorse that regime. (As we discussed above in the context of Rousseau’s general will, the regime must reasonably well serve citizens’ critical interests in order that citizens come to believe that it serves their critical interests.) An additional condition, though, must be appreciable liberty of conscience and expression such that efforts in progressive political theology may take place without censorship or suppression. The Catholic example demonstrates well the necessity of intellectual work internal to a religious tradition that establishes the compatibility of impartiality and religious belief. This is, in fact, the kind of theological argument Locke advanced in his “Letter.” For religious believers to accept the harmonization of loyalties, they must be appealed to from within the perspective of their religious beliefs. Daniel Philpott has argued that religious believers tend to eschew political violence and embrace liberal political orders when their theological systems are capable of affirming political programs of “consensual differentiation.”

Establishing this possibility will take different forms in each religious context, so there is no way to specify in general or in advance what they must look like. In the Catholic case, the theological arguments of John Courtney Murray and Jacques Maritain were hugely influential in

demonstrating that fidelity to a natural law conception of religious truth is compatible with political arrangements that permit freedom of belief – and even “error.”  

119 Without arguments from perspectives indigenous to Catholic theology, neither the teaching authority of the Church nor ordinary believers could have come to regard religious liberty as fully sanctioned by their religious beliefs. Similar efforts occur within other religious systems, of course. Authors such as Farid Esack and Khaled Abou El Fadl are examples of scholars working to ground toleration of pluralism in Muslim tradition and theology.  

120 It is not my place to comment on the merits of those efforts, other than to note them as indications of how religious traditions can evolve and adapt their political theologies to accommodate liberal norms of toleration. In other words, freedom of thought is a necessary condition for the development of these liberal political theologies, but the sufficient conditions will depend on the internal dynamics of each religious group or tradition.

It is worth taking a moment to consider a related, but bolder, line of argument that one might pursue in defense of the harmonization strategy. One long-standing interest of liberal toleration advocates is to convince religious persons that the very nature of religious belief implies the necessity of religious liberty or toleration. This is, for instance, the gist of Waldron’s influential interpretation of Locke’s “Letter Concerning Toleration.”  

121 More recently, Lucas Swaine’s book The Liberal Conscience makes an effort in this vein, offering several lines of argument for toleration that are designed to appeal to what he calls “theocrats,” or persons who would advocate governance based on a comprehensive conception of the good. One such line of

---


121 Waldron, “Locke: Toleration and the Rationality of Persecution.”
argument suggests that theocrats should be wary of establishing their preferred religious doctrine as politically dominant because they have no way of ensuring that the faith isn’t perverted or corrupted over time, resulting in compulsory adherence to wayward beliefs. This is an interesting claim, though not fully convincing, as the potential for corruption is a problem not only for theocrats, but also for any religious person. How can one ever be certain that her religious beliefs have not gone off the rails? This argument undermines not only theocracy, but belief itself. Swaine suggests that religious liberty and democratic contestation are the best bulwarks against corrupted religion, but it is hard to imagine a theocrat agreeing. Rather, a theocrat would likely prefer some system of checks and balances internal to the community of religious authorities, which could ward off doctrinal perversion without opening the door to any and all religious ideas.

A second of Swaine’s arguments suggests that theocrats are rationally committed to the freedom of conscience, both as a descriptive fact (conscience has the capacity to be free) and as a normative claim (conscience ought to be free). The normative claim is entailed by three premises that Swaine thinks theocrats are bound to endorse: that conscience must be free to 1) reject lesser doctrines, 2) accept the good, and 3) distinguish between good and bad doctrines. For conscience to be free in these ways, Swaine suggests, laws must protect freedom of conscience. But this move relies on a conflation of two different kinds of freedom, which the theocrat would surely reject. Conscience is free in the descriptive sense by virtue of its capacity to voluntarily affirm and reject beliefs. The normative premises, in contrast, point to the external conditions affecting the latitude afforded to conscience to act (i.e., evaluate, accept, and reject religious doctrines) freely. The theocrat is surely committed to the descriptive sense of free conscience,

123 Ibid., 49.
but it does not follow that they must oppose political rules and institutions that are designed to
guide persons toward making the right religious choice. In short, Swaine’s argument remains
susceptible to the same objections that Jonas Proast raised against Locke’s arguments for
toleration: theocrats can consistently affirm freedom of conscience as a descriptive fact while
also endorsing institutions (including political institutions) that steer that conscience in the
proper direction.  

The failure of Swaine’s arguments does not rule out the possibility of some successful
defense of toleration and religious liberty from within the logic of theocracy. But it suggests that
we have not yet found it. This would be troublesome if such an argument were the only way for
committed religious believers to endorse liberal toleration. The harmonization of loyalties
strategy, though, suggests that norms of toleration can also arise from within the specific content
of particular religious doctrines, rather than from the structural logic of religious belief itself. We
need not establish that religion as a generic phenomenon entails support for liberal impartiality,
but rather can work to show that specific religions can generate such support.

A further consideration working in favor of the harmonization strategy is
that it may fail by some degree while still providing a great deal of the motivational force that it
would secure if fully successful. To illustrate this, consider a liberal political order in which the
overwhelming majority of persons’ comprehensive doctrines converge on support for the regime.

---

124 Swaine does address Proast’s objection to Locke’s Letter (which Waldron reiterates), which is that the use of coercion is, in practical terms, a reasonable means to draw people to the correct religion “which, without being forced, they would not consider.” Jonas Proast, The Argument of the Letter Concerning Toleration, Briefly Consider’d and Anser’d (Oxford: George West and Henry Clements, 1690), 7. Swaine claims that this is false because the theocrat must be committed to the “principle of rejection”: citizens must be free to reject untrue doctrines, or else they “will be placed at serious risk of being coerced into accepting ungodly practices, which would be a self-defeating outcome on the religious adherent’s very grounds. This, though, overlooks the role of religious authority in the same way that the argument Swaine’s argument from religious corruption does: given reasonable confidence in some religious authority or authorities, the theocrat need not fear the imposition of “ungodly practices,” and thereby need not insist on a political right to reject religious doctrine.
Despite this majority support, there remains a two-percent minority whose religious beliefs reject the general principles of liberal impartiality. That these two percent lack the motivational force supplied by their comprehensive doctrine does not imply a collapse of the harmonization strategy as a whole. For one, in a society that meets the conditions specified in Chapter 3, these citizens will still have various objective reasons to support the regime: it serves their interests, it treats them equitably, it does not disfavor or suppress their religious beliefs, and so on. Their recognition of these reasons may produce some cognitive tension with their religious beliefs, but this does not imply that they will be unable or necessarily unwilling to live in compliance with the political order.

Moreover, that some few citizens do not achieve a harmonization of religious and political loyalties will not necessarily lead other citizens to abandon the beliefs that lead them to support liberal impartiality. Their beliefs may very plausibly remain what they are. Thus, while the regime may lack the degree of stability secured by an overlapping consensus including all citizens, its stability may be weakened by only a very small degree by the failure of some citizens to join the consensus. This is especially easy to imagine in the case of separatist religious groups (such as the Amish or Hasidic Jewish communities) who may be lukewarm in their support for the regime, but who do not actively work to undermine it. In contrast to the civic identity and political emotions approaches, then, the harmonization of loyalties strategy admits of partial successes which neither necessarily nor probably devolve into outright failures.

A final observation about the potential for failure of the harmonization strategy is that some people who appear to reject impartiality toward religious belief do so on the basis of an even deeper commitment to impartiality. Early twentieth-century skepticism toward Catholic political leaders in the U.S. exemplify this well: non-Catholics worried that a Catholic president
would prioritize obedience to the Vatican over loyalty to the American public. While their skepticism may thus appear to reflect a desire to target some persons for disadvantage on the basis of their religious belief (and may have been influenced by xenophobic animus), the substance of their concern also reflected protectiveness of political impartiality. Catholics were perceived to be untrustworthy political leaders because of their justificatory partiality toward the directives of Catholic leaders. Similarly, when Republican presidential candidate Ben Carson said on *Meet the Press* that he was uncomfortable with the idea of a Muslim President, his reason for this overtly discriminatory stance was that he judged Islam to be “[in]consistent with the Constitution.” The problem with a Muslim as President, for Carson, is not that Islam is in some way inherently faulty as a religious system (though Carson may indeed believe that it is), but rather that an adherent of Islam would face too great a conflict of loyalties. They would be unable, Carson worries, to fulfill the responsibilities invested constitutionally in the presidential office. What this suggests is that, even in jettisoning impartiality with respect to religious identity, Carson is actually concerned with maintaining the harmonization of loyalties in elected office, in a manner not dissimilar to historical distrust of Catholic political leadership. By noting these examples, I do not mean in the least to excuse prejudice toward religious minorities, but rather to note that even some positions that violate a principle of religious neutrality do so in a manner that reflects assumptions shared with the harmonization of loyalties approach.

The fundamental virtue of the harmonization of loyalties approach to stability is that it looks to the fact of pluralism as not only a source of challenges for stability (which it is, as we explored in the last chapter), but also as a source of support for stability. Comprehensive doctrines that endorse or are compatible with political norms of impartiality can generate the

---

motivation that is necessary for citizens to reliably uphold impartialist regimes and render them stable. It does this by rooting support in the beliefs and commitments that citizens hold most dear – even if these are not identical for each citizen. The harmonization strategy is not guaranteed to succeed in all settings, but it is a strategy worth pursuing, because it can be realized under reasonably ordinary conditions, and because even partial realizations of the strategy will be beneficial to stability. Impartiality is a sheer face; a pluralism of comprehensive doctrines provides citizens with footholds on which they can maintain their allegiance. In this respect, pluralism is not only perilous, as Hobbes observed, but also promising.

4.6 Moral Communities as a Source of Stability

The argument advanced in this chapter has attempted to show how citizens’ commitments to their comprehensive doctrines can be a source of stability. Because persons tend to hold their religious beliefs very strongly, these beliefs can help to overcome the motivational deficit of liberal impartiality when they support citizens’ compliance with their political regime. It is worth taking a moment to situate my argument in relation to some other claims about the role of religion in shaping citizens and supporting liberal political orders. Doing so should help clarify exactly what I am and am not claiming.

First, what I am defending is not a neo-Durkheimian reliance on religious belief to underwrite social cohesion by sacralizing shared norms. Among other differences, this kind of move is of limited tenability in a context of religious pluralism. I have explicitly assumed that religious beliefs in a pluralistic context root citizens in diverse and, in some cases, incompatible moral and epistemic commitments. Instead, the idea of the overlapping consensus describes a case in which citizens’ diverse conceptions of the sacred point toward the same practical norm:

---

support for the political regime. The beliefs underwriting that norm, and the precise form it takes, may differ widely across religious outlooks.

Second, the argument is related to, though distinct from, views that valorize religion for the civic and political virtues that voluntary, “mediating” associations are purported to instill in citizens. These views view religion primarily as an element of the broader category known as ‘civil society.’ The classic statement in this tradition comes from Tocqueville, who saw voluntary associations as “great free schools” for learning the virtues of democratic life,127 and has found its most prominent contemporary formulation in Robert Putnam’s celebration of choirs, bowling clubs, and other civic associations as generators of social capital and civic engagement.128 Nancy Rosenblum distinguishes two versions of this approach. The first relies on a congruence between the internal character of voluntary associations and the norms of liberal political society, such that participation in liberal-democratic associations will habituate citizens to liberal-democratic norms that will bolster their allegiance to liberal-democratic political societies. The second is less concerned with the character of associations than with their simple existence, as thinkers like Putnam take associations to instill norms of cooperation – a virtue which will spill over to the political sphere.129 My argument, in contrast, is not concerned with behaviors or habits, but rather with belief. As will become clear momentarily, the stability approach takes religious associations to be of paramount importance. But what is significant about religion on my view is that it encompasses moral commitments that individuals hold strongly, and that can outweigh considerations that lean in competing directions, not that it encourages any particular behavioral norms.

127 Alexis de Tocqueville, Democracy in America, ed. J. P Mayer (New York: Harper & Row, 1969), 522. Also see pg. 63: “Local institutions are to liberty what primary schools are to science; they put it within the people’s reach…”
My argument does share a couple things in common with these other approaches. One is an instrumental valuation of religion. From the vantage of stability, religion is not of intrinsic importance, but is rather valuable for its capacity to motivate citizens’ reliable support for a regime. Second, my argument shares in the desire to resist skepticism about the compatibility of religious belief with liberal political order. In the last chapter, we attended to the ways in which religious pluralism does indeed pose a challenge to stability, but in this chapter I have tried to show that it also offers a resource to overcome the motivational deficit of liberal impartiality. What this implies is that if citizens were to lose their attachments to religious beliefs, political stability could suffer. The possibility that religious commitment could wane has been both celebrated and lamented over the history of liberalism. The argument here suggests that, whatever the other merits and demerits of secularization, a diminishment of religious commitment would entail a diminishment of a distinctive resource for political stability.

One of the conditions for belief is the availability of religious and moral communities in which persons learn and come to be committed to their beliefs. The health of these communities is a necessary condition for citizens having the kinds of commitments that can be supportive of stability. This claim may seem counter-intuitive, as religious belief is often cast exclusively in terms of individual conviction: Martin Luther declaring at the Diet of Worms that “Here I stand; I can do no other.” But the characterization of belief as arising solely from the atomistic individual is inconsistent with both historical and contemporary understandings of belief. Rob Vischer, in his book *Conscience and the Common Good*, traces the concept of conscience from its Roman origins through the present day, as treated by such authors as Charles Taylor, and concludes that conscience is intrinsically relational.130 Conscience is intersubjectively formed

---

and, as Taylor notes, identity itself is a matter of relationality: “[the] self can never be described without reference to those who surround it.”\textsuperscript{131} What this means is that belief – or conscience – cannot flourish independent of the flourishing of the communities and institutions that nourish it.

Religious communities, Vischer writes, are “inherently moral enterprises that facilitate the articulation and stabilization of their members’ moral convictions.”\textsuperscript{132} It is by virtue of participating in these two functions that citizens hold the beliefs that they do. Each is of distinct importance. First, religious communities perform a formative and educative function, transmitting beliefs to their members and shaping them in conformity with the values and norms of a tradition. This need not imply stasis or perfect internal homogeneity – part of the formative process may include debate and discernment over values. Second, religious communities sustain beliefs by providing a context in which to express and reaffirm those beliefs. This includes activities of religious exercise: one function of religious rituals is to reinscribe the beliefs and values of a community in its members. By these two mechanisms, religious communities instill and maintain their members’ deep religious commitments.

This discussion is too cursory to definitively establish the precise role of religious communities in developing and sustaining the belief commitments of their members. A full account would need to identify with greater specificity the mechanisms of transmission and support, and would need to determine which precise characteristics of religious communities are necessary and/or sufficient conditions for the performance of these functions. The cursory account above, however, can suffice to ground some preliminary conclusions for law and public policy. Insofar as deep religious convictions are a mechanism for overcoming the motivational


\textsuperscript{132} \textit{Conscience and the Common Good}, 97. The quoted phrase is actually applied to various other associations other than religious ones. We will address the degree to which the argument from stability applies to non-religious associations in the next chapter.
deficit in liberal impartiality, and insofar as healthy religious communities are a necessary precondition for the development of these convictions, then the health of these communities is important for stability. As such, the health of these communities warrants a presumption of support, including by political means as appropriate. Vischer writes, “The viability of conscience will often depend on the viability of the relationships through which conscience is formed. In other words, if our law is to reflect our concern for conscience, it must also reflect a concern for the groups that embody and facilitate individuals’ shared moral commitments.”133 His reasons for valuing conscience are other than stability, but the logic applies nevertheless. If we hope to find a resource for stability in religious belief, then it is necessary to support the communities which are a precondition for religious belief.

Determining exactly what this support entails in political terms requires a lengthier discussion than is possible here. The next chapter attempts to work out some of the implications in relation to the narrow question of religious exemptions. What we can say in general is that the harmonization of loyalties strategy depends upon the flourishing of religious communities at least to the extent that they are able to perform the belief-inculcating functions we have discussed. While it is beyond our capacity to identify all of the ingredients of flourishing for a religious community (and to determine which ingredients are necessary for the harmonization of loyalties strategy to succeed), we can identify some requirements in negative fashion. For one, religious associations must not be overtly suppressed: a religious community that is not permitted to meet or to give voice to its beliefs will struggle to thrive in the ways necessary to transmit and sustain its distinctive body of beliefs. So, religious associations must be afforded the basic liberties of association, assembly, and exercise.

133 Ibid., 71.
The exact limits of these liberties will have to be established by argument, but it is clear that support for religious associations will need to go beyond merely formal guarantee of certain rights. Religious associations and citizens alike may struggle to maintain their distinctive belief commitments in the face of political and legal frameworks that, while not overtly aimed at suppressing religious belief, nevertheless contribute to the development of a “naked public square” in which religious belief is marginalized and/or relegated exclusively to the private square. This outcome might be lamentable for any number of reasons, but the relevant one here is that it would strip away a resource for motivating support of liberal political institutions. Devising laws and institutions that not only permit the existence of religious associations but moreover support their thriving can not only respect rights and individual human dignity, but can also help sustain an important source of stability. In the next chapter, we turn to the practical question of what kinds of accommodations are conducive to the flourishing of religious communities toward the end of maintaining religious beliefs for the sake of political stability.

---

CHAPTER FIVE:

STABILITY AND THE ACCOMMODATION OF RELIGIOUS BELIEF

We are, at last, prepared to return to the issue of religious exemptions. Though much of the intervening argument has pertained to issues of pluralism broader than the specific matter of exemptions, it has been with an eye toward addressing this particular problem for contemporary liberal societies that our investigation has proceeded. Religious exemptions pit the commitments of religious liberty and equal status under the law against one another, and have generated great controversy and consternation. In the language of this project, they represent a deviation from liberal impartiality. Our goal here is to formulate a normative framework for exemptions on the basis not of equality (as in the approaches we considered in chapter 1), nor any other substantive moral commitment, but rather of stability. Rawls’s conception of stability for the right reasons provided a normative ideal, from which we derived principles of distributive and justificatory impartiality in chapter 3. In chapter 4 we amended that account to recognize some ways in which a pluralism of comprehensive doctrines could serve the stability of an impartialist political regime. Here, the argument turns more specific, attempting to answer the basic question motivating this project: when, if ever, does a religious belief or religious identity warrant a dispensation from legal obligations?

Before proceeding, a few methodological notes are in order.

First, in this chapter, as in other parts of the argument, I will speak of both exemptions and exemption requests. The two terms serve to differentiate between exemptions that have been granted and those that are under consideration. In most of what follows, however, the terms are often used interchangeably. Given that this is an exercise in abstract, largely hypothetical theorizing, exemptions and potential (i.e., requested) exemptions enjoy much the same status.
Second, I do not distinguish judicial exemptions from legislative or administrative exemptions. Whether an exemption is granted (or denied) by a judge, in a post hoc evaluation of a law, or by a legislature in the text of a law, or by an executive or administrative agency working out the application of the law, the philosophical issues are basically the same. In practice, these various sites of exempting authority may address distinct questions and may yield distinct results, but our investigation is one of idealized regime-design, so we will ignore these differences and their practical implications, whatever they might be.¹

Third, I proceed on the assumption that all religious exemption requests operate on the same basic logical structure. The basic logic of a religious exemption is that it strives to resolve a choice that some individual faces between violating a religious commitment and violating some legal requirement. I resist the temptation to differentiate exemptions along divides such as whether they provide some positive benefit versus relief from an obligation, or whether it is a positive obligation versus a negative restriction. Whether an exemption involves doing something that is (religiously or legally) prohibited, failing to do something that is (religiously or legally) required, and so on, is ancillary to the basic problem. The law establishes some baseline of obligation, and the question for religious exemptions is whether to alter the obligations for some persons on the basis of their religious beliefs or identity.

Fourth, this chapter includes discussion of a number of legal cases involving religious exemptions. I involve these for two reasons. First, in a discourse that is largely abstract, the cases provide some concrete specifics that both illustrate the significance of abstract concepts and illuminate the very real stakes of disputes over religious accommodation. Second, in recent years, a number of cases regarding religious exemptions cases have assumed a prominent spot in

¹ Recall (from footnote 14 in the Introduction) that an exemption is a dispensation from a legal obligation that does not imply any alteration of the content or general applicability of those obligations.
American political discourse. Insofar these are of contemporary interest, they merit our consideration and provide a proximate occasion to observe the practical implications of the argument from stability.

The chapter proceeds in several parts. In the first section, I draw on the arguments of preceding chapters to identify in general terms the normative implications of the end of stability for the right reasons for the question of religious exemptions. In section 5.2, I identify several principles that a religious exemption framework ought to respect in order to maximize the stability-enhancing benefits of exemptions and minimize their potential threat to stability. In 5.3, I provide a brief overview of U.S. jurisprudence regarding exemptions, and summarize the current legal paradigm governing them, as supplied by the Religious Freedom Restoration Act (RFRA). Then, in section 5.4, I turn to the most prominent recent U.S. case involving religious exemptions: the request by Hobby Lobby stores for an exemption from the contraception coverage mandate in the Affordable Care Act (or Obamacare). I consider how the stability approach would judge Hobby Lobby’s request, and relate this evaluation to the ruling provided by the Supreme Court. Section 5.5 offers a comparison between the stability framework and the framework established by RFRA, and then identifies the merits of the stability approach in this comparison. Finally, in section 5.6, I consider four potential objections to the normative approach I have advanced here. This concludes the substantive portion of the present project.

5.1 The Ambivalence of Stability

On the basis of the arguments advanced in chapters 3 and 4, we are in a position to advance some general claims about the implications of stability for religious exemptions. While a categorical answer to the exemption question – a blanket yes or a blanket no – would be

conceptually neat and practically straightforward, the arguments so far do not support an unequivocal stance. Rather, the end of stability is thoroughly ambivalent on whether exemptions from generally-applicable laws on the basis of religious belief are permissible, or perhaps even necessary. In this section, I want to advance three distinct claims, each derived from considerations raised in the preceding chapters. One of these claims weighs against the permissibility of exemptions; the other two suggestions reasons why exemptions may be desirable, at least in some cases. The tension created among these three claims will provide the normative parameters for the next section, in which we specify practical guidelines for granting or denying certain exemptions.

The first claim is as follows:

**Thesis 1**: The norm of distributive impartiality, which is a requirement of stability, implies a presumption against religious exemptions.

The third chapter established that stability requires two general norms of impartiality – distributive and justificatory. Regarding the former, I argued that stability requires a general norm of uniform legal treatment: individuals ought to be subject to uniform legal obligations, determined without respect to certain of their individual attributes. This is a presumptive norm, but not an indefeasible one, of course, as some attributes (level of income, mental health status, e.g.) can reasonably be regarded as relevant to the political distribution of burdens and benefits. But the general presumption is against consideration of individual characteristics in the determination of that distribution. Religious belief and religious identity were counted among the
characteristics categorized as presumptively politically irrelevant. The political order should be impartial with respect to citizens’ religious commitments.

An exemption is inherently an instance of partiality. In respect of some religious belief, members of the class of persons sharing that belief are granted a set of legal obligations that is altered to differ from the default set of obligations borne by other members of the political society. Religious exemptions recognize the religious beliefs of some persons as relevant considerations in determining their legal obligations. Take the mechanism of conscientious objection in the U.S. as an example. During times when the draft is active, male citizens of a determined age are required to present themselves for military service. Male citizens of that specified age whose religious beliefs oppose the use of force in war may be excused, however, from that requirement. Many other considerations – fear of death, squeamishness, indifference to the American cause, and so on – may make a person averse to military service, but the law does not consider them grounds for dispensation. The law of conscientious objection thus demonstrates partiality with respect to religious beliefs that stand opposed to war.³

The argument in chapter 3 did not categorically preclude consideration of religious belief and/or identity in the distribution of political benefits and burdens, but it did establish a strong presumption against such consideration. While the Hobbesian argument from stability does not definitively rule out religious exemptions, it at least puts exemptions under a heavy burden of justification. A second claim supplies one of the considerations that might override the presumption against exemptions.

³ Not insignificantly, U.S. conscientious objector laws also permit nonreligious philosophical exemptions.
**Thesis 2**: Some exemptions from generally-applicable laws may be warranted to ease substantial burdens (on religious exercise) that burdened citizens cannot regard as justified.

Recall that the conception of stability we have adopted as a normative ideal – stability “for the right reasons” – rests on citizens’ support for their regime motivated by endorsement from within the perspective of their conception of the good. A regime that is stable for the right reasons is one that citizens regard as justified from their particular moral and epistemic vantage. What this implies, primarily, is that regimes should be of the sort that can win endorsement from a wide variety of comprehensive doctrines. The argument in chapter 3 established that norms of impartiality will ordinarily be best able to secure support from citizens of diverse stripes. But there are occasions in which distributive impartiality will actually make it difficult for some citizens to endorse their regime.

The reason is that uniform treatment under the law does not necessarily entail uniform consequences for diverse citizens. Given differences in citizens’ preferences, projects, and commitments, laws of general applicability may impose unequal burdens, from a subjective perspective, on different individuals. For instance, gun control regulations are of different consequence for a pacifist who shuns tools of violence, on one hand, and a devoted hunter who understands that hobby as central to his identity, on the other. This dynamic inspires an interesting debate over whether this uneven distribution of burdens is consistent with a normative commitment to equality – we considered this question briefly in section 1.5, with Bhikhu Parekh and Brian Barry taking opposing views. Our topic here is not how to secure equality, however, but rather stability. From the stability vantage, the uneven distribution of laws’ burdens is indeed
of significance, for a legal order that severely restricts a citizens’ ability to live in accordance with her deepest commitment is unlikely to win her endorsement – even if this restriction is not the result of (unfriendly) partiality toward that person.

Recall Gaus’s characterization of the kinds of regimes that could become the object of a convergence of citizens’ reasons for endorsement. He wrote that no such convergence would be possible around regimes that “build social order on the requirement that some renounce their most cherished convictions” or “abandon our deep ideals as the basis of our own existence.”

Consider, for instance, the cases of sacramental substances mentioned above. Most Americans do not consume peyote, and many of those who do have no deep attachment to it – their interest is recreational, and they could presumably achieve the same ends reasonably well with some substitute drug. A ban on the possession and consumption of peyote does not pose a significant burden – from a subjective vantage – on these people. For members of the Native American Church, however, a prohibition on peyote makes participation in a core religious practice illegal. Citizens in this situation are thus faced with competing obligations – to their religious commitments and to the requirements of the law. In order to be compliant citizens, they must forgo an important aspect of their religious identity. This conflict provides citizens with a reason against endorsing their regime, and may indeed override whatever reasons they have for endorsing it.

Of course, to live under a system of civil order is to endure limitations on actions that one would like to take. This includes actions motivated by religious belief or commitment. An effective rule of law will inevitably preclude or hinder some instances of religious exercise. Very often, these are not injustices or dilemmas, but rather mundane facts of life. Often, these burdens

---

are unremarkable because the restriction they impose is minimal. For instance, the installation of bicycle lanes on the street in front of a church “burdens” the individuals who would otherwise park there while attending services, but it is a minimal burden: they can park around the block and conduct their religious activities after a short walk.\(^6\) I assume that these minor constraints on religious practice do not appreciably undermine most persons’ support for their regime, as their ability to fulfill both their religious and their civic obligations is only minimally impeded. What concerns us, in regard to stability, are legal restrictions that \textit{substantially} burden religious exercise. (We will return in more detail to the question of how to determine whether a burden is substantial below.)

Even among laws that substantially burden religious exercise, though, there are many such restrictions that citizens tolerate day-in and day-out, without much notice. For instance, the Gospel of Matthew reports that Jesus instructed his followers to “make no oath at all” (Matthew 5:34). Modern systems of contract and litigation, however, depend on the regular swearing of oaths. No person could expect to get by without doing so (at least implicitly). What makes burdens such as this tolerable is the fact that they are easily regarded as justifiable on the basis of some other compelling interest. The burden imposed on Christians who are required to make oaths in violation of their beliefs is justified in light of the easily-recognized bads that the requirement avoids: fraud and manipulation, namely.\(^7\) This means that some burdens on religious exercise are not undermining to stability, because the persons burdened can themselves regard


\(^{7}\) It is also the case that religious persons and communities often interpret their traditions so as to resolve apparent conflicts with non-religious obligations. So, many Christians do not regard mandatory oath-swearing as a burden because they discount, disregard, or interpret away the apparent scriptural prohibition.
the burden as justified from the perspective of their conception of the good.\(^8\) This is the case even when some citizen may not directly endorse the end served by the burden-generating law, so long as she regards it as a reasonable end for others to pursue. For instance, a smoker might prefer, all else being equal, to be allowed to smoke in a restaurant, but can recognize the public value of prohibiting indoor smoking. We live always under myriad legal burdens, and even when these burdens impose some appreciable cost to pursuing one’s good in one’s own way, we regard a great many of them as fully justified.

There are instances, however, where a citizen cannot regard a burden on his or her religious exercise as justified. When these burdens are significant, they will amount to the kind of infringement on citizens’ pursuit of their good that undermine those citizens’ ability to endorse their regime, consequently undermining stability. But what does it mean for a citizen to be unable to regard a burden on their religious exercise as justified? There are at least three situations that could produce this result.

First, if the application of some burden to a citizen does not serve a compelling public interest, it will be difficult for that citizen to regard the burden as justified. Requirements that citizens make and keep oaths, prohibitions on indoor smoking, and many other burdensome laws are easily regarded as justified because of the clear public interest they serve. To use the language of contemporary theory, they are supported by freestanding, public reasons. Even when a particular citizen does not personally endorse the regulation, they can access and appreciate the reasons why others do. Absent a publicly accessible justifying reason, however, burden-imposing

---

\(^8\) This assumes that persons are capable of maintaining their conception of the good even when they recognize that it cannot be perfectly instantiated. An alternative view might hold that persons abandon or revise their conceptions of the good when they find that it includes contradictory elements. This does not seem true, however, as persons regularly accept limitations imposed by their conception of the good on their own pursuit of some aspect of their good as they understand it. Significant discord among elements of a conception of the good may prompt players to revise their commitments, but persons are not absolutely intolerant of discord.
laws may fail to be justified to the citizens whose religious exercise they restrict. The 2005 case of *Gonzales v. O Centro* illustrates this well.\(^9\) The case concerned a New Mexico branch of a Brazilian religious organization whose members desired to import, distribute, possess, and consume *hoasca*, a hallucinogenic tea, for sacramental purposes. *Hoasca* was (and is) prohibited under the Controlled Substances Act, and the case asked whether O Centro was owed an exemption. The Court ruled unanimously against the government on grounds that applying the burden of the law to the members of the organization did not advance a compelling governmental interest. This is not to say that the law *itself* advances no interest, but rather that *applying the law* to the individuals in question does not. The Court’s ruling did not turn on the justifiability of that burden, of course, but the case illustrates the role of a compelling interest in rendering a burdensome law justifiable. When some citizen’s religious exercise is restricted for no good reason, the likelihood that she will be able to regard that restriction as justified decreases.

Second, a burden is less likely to be regarded as justified when the restriction it imposes on religious exercise goes beyond what is necessary to achieve the end it is intended to serve. Persons are capable of recognizing some restrictions on their actions as justified in light of some compelling interest, but the presence of such an interest is not a blank check to impose any restriction whatsoever. Burdens that exceed the measures necessary for the end they are intended to serve are less likely to be regarded as justified from the perspective of those burdened. An example of excessive burdening is found in the 1993 case of *Church of the Lukumi Babalu Aye v. City of Hialeah*, in which the Supreme Court found unconstitutional a city ordinance prohibiting

---

ceremonial killing of animals.\textsuperscript{10} The ordinance was very clearly designed to target practitioners of Santeria, whose rituals sometimes involve animal sacrifice.

The Court ruled that the prohibition on animal sacrifice was broader than necessary to achieve the ends that its defenders claimed it was intended to serve. Justice Kennedy, wrote, for the Court, that “The legitimate governmental interests in protecting the public health and preventing cruelty to animals could be addressed by restrictions stopping far short of a flat prohibition of all Santeria sacrificial practice.”\textsuperscript{11} If the city legislators who wrote the ordinance were truly interested in doing what they said they were, they could have achieved their ends without so severely restricting religious exercise. Instead, their attempt to target animal sacrifices revealed the true intentions motivating the law: animus toward practitioners of Santeria and a desire to inhibit their religious exercise.\textsuperscript{12} Laws that go beyond what is necessary for the satisfaction of a public interest raise suspicion of discriminatory intent, and are less likely to be regarded as justified from the perspective of those whom they burden.

Third, there are some instances in which a citizen affirms a system of belief that precludes endorsement of a burden-generating law even when it is supported by a suitably public reason. This is the most common variety of conflict between religious belief and public obligation. It can be observed in the spate of recent cases concerning the legal recognition of same-sex marriages and the attendant obligations on employers, public officials, or business-owners.\textsuperscript{13} These cases arise from a conflict between legislative or judicial decisions enacted on

\textsuperscript{11} Ibid., 508:538.
\textsuperscript{12} This insight is supported by the legislative history, including meetings of the Hialeah city council, where the council president asked “What can we do to prevent the Church from opening?” Ibid., 508:541.
the basis of constitutional goods such as equal protection or substantive due process, on one hand, and religious views that cannot regard granting same-sex couples the full status of marriage as a public good. Moreover, they cannot regard as justified any requirement that individuals who oppose same-sex marriage personally recognize and/or facilitate those unions. The issue is not that there is no compelling interest behind the law, nor that the law goes beyond what is necessary to satisfy that interest, but rather that some religious beliefs directly oppose same-sex coupling, and therefore cannot regard any burdens imposed by recognition of same-sex marriage as justified. The same dynamic is at work in many cases of religious conscientious objection to military service: guaranteeing the national defense is an end supported by fully public reasons, yet committed pacifists are unable from the perspective of their own comprehensive doctrines, to regard the end, and its attendant obligations, as justified.

In cases falling into any of these three categories, exemptions may be useful for relieving burdens that citizens cannot regard as justified, and thereby restoring their likelihood of endorsing and supporting their regime. Note that this standard does not sanction exemptions for any and all objectors to legal restrictions. Citizens must legitimately be unable to regard their burden as unjustified. For instance, persons who object to the requirements of the law simply because they dislike the material imposition would not be included. Persons who object to legal requirements on irrational or unsound bases would be similarly excluded. For an example of the latter, note that the plaintiffs in Burwell v. Hobby Lobby requested an exemption from a requirement to provide health insurance coverage for some forms of contraception they believed to be abortifacients. There is some significant doubt regarding the factual accuracy of that

---

belief. The government did not pursue that line of argument in court, but if it were to be shown that Hobby Lobby’s objection to the legal burden rested on faulty information or faulty reasoning, the above rationale for granting their exemption would be inapplicable.

Thus far, I have argued that (1) stability requires a general norm against exemptions, but that (2) exemptions may be warranted to remedy burdens that citizens cannot regard as justified. Note that this could include group exemptions (such as in O Centro) for the purpose of relieving burdens on individuals’ religious exercise. This is the case when the communal religious practice is a necessary condition for the individual’s successful exercise of her religious beliefs.

There is a third normative claim that I wish to advance, deriving from the discussion of motivational efficacy in chapter 4. There, I argued that a pluralism of moral communities can indirectly contribute to stability by fostering the deeply-held beliefs that can motivate citizens to maintain their support for a regime in the face of temptations to non-comply. This leads us to the third claim:

**Thesis 3**: Some exemptions enhance stability indirectly by supporting the flourishing of diverse moral communities that provide a necessary condition for effective motivation to support a regime.

The practices of religious exercise that may come into conflict with requirements of the law pertain not only to individuals, but also to collectives. Our second thesis, above, may supply justification for exemptions in some of these cases, insofar as burdens on collective religious exercise threaten citizen’s support for the regime. But the claim from the distinctive usefulness

---

of moral community adds a further basis for exemptions for communal religious practice.

Restrictions on individual exercise poses a conflict for those individuals (obey either their religious dictates or the law), but restrictions on group exercise pose a distinct threat to religious identity. Namely, restricting communal religious practice threatens to erode the distinctiveness of that moral community itself.

Two examples serve to make the point. First, just as some key conflicts in recent years have concerned the use of prohibited substances for sacramental purposes (Employment Div. v. Smith, Gonzales v. O Centro, e.g.), a conflict over a banned substance arose nearly a century ago in the context of Prohibition. A blanket ban on the possession and consumption of alcohol would have posed a great challenge to the American Catholic community, for whom wine is sacramentally important. In that instance, an exemption for sacramental use of wine was written into the National Prohibition Act. Absent such an exemption, the central sacramental ritual for Catholic citizens would have been significantly constrained. In consequence, the distinctiveness of Catholic religious exercise and identity would have been diminished. For a second example, consider what is known as the “ministerial exception” to the Civil Rights Act of 1964, which permits churches and religious institutions to discriminate on the basis of religion in hiring and firing their ministers. If not for this exemption, the state would be able to establish criteria for the hiring and firing of ministers, as it can for other employment contexts. The Civil Rights Act’s prohibition on sex discrimination could threaten the Catholic church’s all-male priesthood, forcing ministerial hires to accept women candidates on the same footing as men. This would, to say the least, create difficulties for religious communities in maintaining their distinctive identity and charism.

---

Each of these examples serves to demonstrate that some generally-applicable laws, when they burden collective religious practice, have the potential to erode the identity of those religious groups. The ultimate concern in this case is not simply that some individuals are torn between their religious and their legal obligations, but that the very *source* of their religious commitments might be undermined. Moral communities are not impervious to suppression by the political order – even when this suppression is unintentional. In fact, if a plurality of moral communities is to flourish, the political order must deliberately create or at least permit social space for this to occur. In the last chapter, I suggested that this means that stability leans in favor of ample freedom for religious individuals to hold, express, and act on their beliefs, and, moreover, latitude for religious communities and institutions to enact shared ritual practices and ways of life. It follows, then, that exemptions to protect communal practice can in this way be indirectly supportive of stability.

The claims advanced in this section amount to an application of the arguments in chapters 3 and 4 to the question of religious exemptions. Stability is ultimately ambivalent on exemptions, in that it both supports a presumption against differential treatment under the law and yet recognizes two ways in which exemptions can support stability. A general norm of impartiality in the face of pluralism is conducive to stability, but pluralism also means that diverse persons will be burdened differently by general laws, and therefore securing the endorsement of some may require deviations from the impartiality. Stability thus leans both for and against exemptions. It is important to reiterate that the second and third claims do not provide blanket justification for any and all exemptions, but rather supply two distinct justifications for *some kinds* of exemptions. The task that remains is to further identify the parameters for which
exemptions might be warranted as deviations from the norm of impartiality that, on balance, enhance the stability of a political order.

5.2 Some Guiding Principles

For an account of religious accommodation that takes stability as its guiding end, the goal must be to identify those categories of exemptions whose stability-supporting effects outweigh their stability-undermining effects. We might illustrate this dynamic by the following graphic:

![Graph showing the relationship between stability and permissiveness of exemptions]

From the baseline of total distributive impartiality with respect to religious belief (point A in the figure below), some strategic exemptions can result in an increase in stability. But the relationship between religious accommodation and political stability is not linear. Rather, it must reach some maximum (B), beyond which point further deviations from distributive impartiality will result in decreased stability (C). What we want to determine are principles for identifying exemptions that will result in a net increase in the stability of the political order.

The figure is misleading in at least the following way: it suggests that “permissiveness” is a single-dimensional variable, when in fact it is multi-dimensional (as the following discussion
will make clear). Nevertheless, it serves to illustrate in basic terms the relationship between stability and exemptions by collapsing these dimensions into a single characteristic of “permissiveness.”

More importantly, though, it reflects an assumption that I have not defended: that there exists a “virtuous mean” (point $B$) between the extremes of no accommodation and total permissiveness, the achievement of which will maximize stability. All I have established so far is that stability is ambivalent on exemptions. It is conceptually possible that, from the vantage of stability, either extreme approach to exemptions will be preferable to some middle ground. So why assume that stability is a matter of some exemptions, rather than none or all (or, less absurdly, many)?

I will not attempt here to prove in conceptual terms the existence of the virtuous mean that I have assumed above. But I can say two things in defense of moving forward in this way. First, it is the middle position that best reflects common intuitions regarding accommodation of religious difference in liberal societies. Rejecting all exemptions out of hand is unnecessarily harsh, but there must be limits to what religious belief may justify. Now, this statement is something of a cheat, since it goes beyond what I have defended on the basis of stability alone. That said, I do not mean to place too much weight on it. My point is only that insofar as the stability-based approach to exemptions is designed to intervene helpfully in contemporary American discourse about religious accommodation, it is best situated in the middle ground occupied my most liberal polities.

Second, and more importantly, the goal of the following sections is to show that it is possible to identify certain categories of exemptions with significant benefits for stability and minimal threats to stability. That is, by distinguishing various kinds of exemptions, we can
identify those that are most net-supportive of stability. The idea, then, is to permit exemptions that are most likely to result in an overall benefit to stability, and to proscribe them when they result in an overall diminishment of stability. In his opinion in the pivotal case of Employment Division v. Smith, Justice Antonin Scalia expressed a desire to “limit the impact” of a presumptive claim to religious exemptions. His concern was that opening the door to exemptions would require them in all corners of public life. He concluded, though, that there is no judicially acceptable means of discriminating among exemption requests, granting some while denying others.\textsuperscript{18} We are not here bound by the same limits as the Court, however. One way of understanding what follows is to identify normative parameters that do what Scalia hoped for. This approach aims to show the existence of some virtuous mean by identifying it in practical terms.

\textit{General vs. Targeted Exemptions}

In U.S. law, there are currently two rubrics under which religious exemptions may be granted: exemptions that specify practices from which persons with religious objections can be exempted, on one hand, and general provisions for anyone with an objection to a law, on the other. The former approach involves identifying particular laws that burden, or are likely to burden, some individuals’ religious practice, and then to grant exemptions from those laws. Examples of exemptions falling into this category include congressional approval of a statutory exemption from military codes of dress for servicemembers whose religious beliefs compel them to wear special headcoverings – such as yarmulkes.\textsuperscript{19} The alternative to granting exemptions “targeted” at particular practices is to establish a general norm of exemptions for the purpose of

\textsuperscript{18} Employment Division v. Smith, 494 U.S. 872, 886 (1990), 494.

\textsuperscript{19} 10 U.S.C. § 774(b)(2) (2012).
accommodating religious exercise.\textsuperscript{20} The Religious Freedom Restoration Act of 1993 (RFRA) is an example of this latter approach: it does not specify which practices are to receive exemptions, but rather grounds a claim by anyone who judges their religious exercise to be burdened by \textit{any} law whatsoever.

The key difference between these two approaches is that, under targeted exemptions, the exempting agent (Congress, a judge, etc.) has a rather clear idea of what conduct is being permitted and what is not,\textsuperscript{21} whereas under general exemption frameworks, the content of the exemptions that may ultimately result are not known. The consequence of this difference can be observed in recent statements from early supporters of RFRA, who endorsed it two decades ago as a protection for practitioners of minority religions, but did not expect it to be used by Christian conservatives, as it has in recent years, to resist progressive developments in social morality, such as increasing recognition of same-sex marriage.\textsuperscript{22} This lends a degree of risk to the general exemption framework: one cannot know in advance which laws will give rise to requests for exemptions, and there is no way to categorically protect particular laws from the possibility of religious exemptions.\textsuperscript{23}

Justice Antonin Scalia, in his \textit{Smith} opinion noted a concern along these lines. That decision (in)famously rejected the possibility that the Free Exercise clause of the First Amendment guaranteed citizens exemptions from neutral laws that burden their religious

---


\textsuperscript{21} Though not a perfect idea, as the codified exemption remains subject to interpretation.

\textsuperscript{22} See \textit{amicus curiae} brief for United States Senators Murray, Baucus, Boxer, Brown, Cantwell, Cardin, Durbin, Feinstein, Harkin, Johnson, Leahy, Levin, Markey, Menendez, Mikulski, Reid, Sanders, Schumer, and Wyden as Amici Curiae In Support Of \textit{Hobby Lobby} Petitioners And \textit{Conestoga} Respondents, \textit{Burwell v. Hobby Lobby Inc.}: “[W]e could not have anticipated, and did not intend, such a broad and unprecedented expansion of RFRA.” Available at: http://www.murray senate.gov/public_cache/files/65b9e452-293c-4897-a415-4341d6166371/hobby-lobby-conestoga-amicus-brief.pdf.

exercise. Justice Scalia’s opinion rested, in part, on a worry that to establish this precedent would lead to a cavalcade of exemption requests, with citizens seeking dispensations from “civic obligations of almost every conceivable kind.”

He even provided a list of potential such requests, which Justice Sandra Day O’Connor, in her concurring opinion, referred to as Scalia’s “parade of horribles.” The list included:

“compulsory military service... payment of taxes... health and safety regulations such as manslaughter and child neglect laws... compulsory vaccination laws... drug laws... traffic laws... social welfare legislation such as minimum wage laws... animal cruelty laws... environmental protection laws... and laws providing for equality of opportunity for the races…”

Scalia’s concern was that if the Court granted one exemption (for consumption of peyote), it would be bound by precedent to grant all these exemptions. This would create a situation bordering on anarchy. Justice O’Connor rightly notes, however, that this is not the case. Courts have done a reasonably good job at striking a balance between competing considerations and rejecting exemption requests when warranted. Indeed, there are any number of bases on which exemption requests might be declined – even given a presumption of the legitimacy of such requests. Admitting the possibility of exemptions from the full range of legal requirements does not mean that exemptions must be granted from any and every legal requirement. To open the door to exemptions is not to lay out a welcome mat for the entire horrible parade.

26 Ibid., 494:872.
27 Ibid., 494:902.
There is, additionally, a positive reason to prefer a general exemption framework over a system of targeted exemptions. A general exemption framework maintains impartiality among religious views by not presuming to know which legal obligations burden religious exercise. A system of targeted exemptions, in contrast, relies on some political authority (whether legislative, executive, or judicial) to identify laws that burden religious practices and grant an exemption. This approach is likely to result in a greater degree of accommodation to adherents of majority religions than to adherents of minority religions, insofar as the former’s commitments will be more reliably reflected in democratic decision-making. For instance, as mentioned above, the use of wine in Catholic sacramental rites was exempted from the ban on alcohol during prohibition, yet the sacramental use of peyote by the Native American Church or hoasca tea by the O Centro community did not receive proactive exemptions from the legislature. Instead, they were forced to press their respective cases through litigation.

My point is not that proactive legislative exemptions are preferable, but rather that a system of targeted exemptions is more likely to advantage members of majority faiths whose views are more familiar to lawmakers and more likely to figure in their deliberations. (This does not apply only to legislatures: evidence shows that courts also tend to be more accommodating to mainstream religious views with which they are familiar. 28) We briefly explored, in section 1.5, Will Kymlicka’s account of multicultural rights, which relied on the observation that members of minority cultures must expend special resources to maintain their cultural identity, as democratic decisions will be less likely to automatically accommodate their commitments than those of persons belonging to majority cultural groups. 29 My point here is similar. Systems of targeted

exemptions are likely to favor members of mainstream religions, thereby violating the norm of impartiality among religious persons and groups. Stability therefore leans in favor of a general framework that establishes the basic legitimacy of exemption requests, rather than a program of targeting particular practices for exemptions from particular laws.

*Accommodative Equity*

Impartiality implies a general presumption against exemptions, but even when this presumption can be overridden, stability still requires impartiality in terms of which exemption requests are granted. There are two dimensions to this application of impartiality. First, exemptions should be granted or denied impartially with respect to particular religions. They should not be granted to adherents of religion $A$ but not to adherents of religion $B$. For instance, all being equal, an exemption for one religious group to consume a prohibited substance for sacramental purposes should be accompanied by a parallel exemption for another religious group to consume a different substance of sacramental importance. Of course, very often all else will not be equal: some features distinguishing the cases will warrant different outcomes. (For instance, one prohibited substance may be appreciably more dangerous or more addictive than another.) But these different outcomes must rest on relevant differences between cases, and not on partiality (negative or positive) toward particular religions or their adherents.

Second, religion should not be disfavored, in comparison to other considerations, as a grounds for an exemption. Where there are exemptions granted for reasons other than religion, exemptions for religious reasons are also warranted. A paradigmatic example of this kind of impartiality is found in a case involving the Newark Police Department’s requirement that
Two Muslim officers challenged the rule on grounds that it restricted their ability to comply with religious requirements to grow beards. A Circuit Court ruled the restriction unconstitutional on grounds that it allowed exemptions for some medical reasons (skin conditions that make shaving difficult) but not religious ones. This policy, the Court wrote, “raises concern because it indicates that the Department has made a value judgment that secular (i.e., medical) motivations for wearing a beard are important enough to overcome its general interest in uniformity but that religious motivations are not.” If exemptions from some law or requirement are granted for non-religious reasons, then the presumption is that religious exemptions should be granted as well.

Note, though, that this equity is a one-way street: religion must be treated *no worse* than non-religious considerations, but the converse does not necessarily hold. The reason is that religious beliefs are, from the vantage of stability, of special importance for the central place they occupy in the lives of believers, and for their ability to undermine or strengthen support for a regime. We will return below (section 5.6) to the distinctiveness of religious belief, but for now we must simply note that the principle of accommodative equity requires that religion not be comparatively disfavored, but not necessarily that non-religious considerations be granted equal accommodation.31

---

30 Fraternal Order of Police Newark Lodge No. 12 v. City of Newark, 170 F.3d 359 (3d Cir. 1999).
31 I leave aside here the challenge for accommodative equity that we discussed in Chapter 1 in relation to Sager & Eisgruber’s “equal liberty” approach: how to determine which non-religious considerations are sufficiently analogous, or of comparable weight, to serve as the benchmark for treatment of religion. See, e.g., Thomas C. Berg, “Can Religious Liberty Be Protected as Equality?,” *Texas Law Review* 85 (2007): 1185–1215. Working out the concrete application this principle requires some estimation of religion’s importance vis-à-vis other concerns. I don’t have any proposal for resolving this problem, so I must concede that some cases will be difficult to adjudicate. But in some cases, such as *Fraternal Order of Police v. Newark*, the norm will be of clearer consequence. This may seem like naked evasion of a conceptual incoherence, and to some degree it is that, but it is also a recognition that even an incompletely theorized principle can be of meaningful practical use.
Sincerity

Turning now to more specific principles for discriminating among exemption requests in order to determine which to grant and which to reject, the first and most basic parameter is that exemptions should be granted only in cases where a law burdens a religious practice or commitment that a citizen holds sincerely. This may seem obvious, but it is worth stating. The reason for this restriction is that if exemptions are admitted simply on the basis of a person’s claim that they hold some belief, persons will have incentives to claim religious beliefs that they do not really affirm in order to skirt the law for other reasons. For instance, in the case of United States v. Quaintance, defendants who had been prosecuted for marijuana possession defended themselves on grounds that as members of the Church of Cognizance (which they founded), they were compelled to consume marijuana, which according to the teachings of the Church “is a deity and a sacrament.” The circuit court rejected their request on grounds that the defendants’ religious beliefs were being proclaimed in pursuit of “commercial or secular motives rather than sincere religious conviction.” In general, religious petitioners who reach advanced stages of litigation pursue their claims on the basis of sincere beliefs, but courts have not hesitated to reject claims on the basis of insincerity.

The sincerity requirement is intuitively sensible, but it merits spelling out the stability-based reasons behind it. First, limiting exemptions to cases involving sincere beliefs ensures that exemptions are granted only to persons for whom the exemption will relieve the kind of

---

32 United States v. Quaintance, 608 F.3d 717, 718 (10th Cir. 2010).
33 Ibid., 608:722.
unjustified burden described in section 5.1. For any law, there are surely many persons who would happily feign belief in order to secure relief from the burden it imposes, but, as explicated above, stability does not require relieving persons of any and all burdens. Second, a sincerity standard helps other citizens recognize the legitimacy of the exemption. When others see that an exemption is granted in order to resolve a sincere conflict, they will be less likely to feel disadvantaged or taken advantage of. Suspicion of insincere manipulation, on the other hand, can generate resentment and compel other citizens to seek their own ways to shirk their legal obligations.

How is sincerity to be assessed? One means is to directly judge the sincerity of a claimant’s testimony. A person seeking a religious exemption can be asked to explain or describe the beliefs motivating their request, and a judge can make an evaluation of the individual’s apparent sincerity, not unlike how a judge would evaluate the credibility of a witness in some other context. This direct mode of assessment will go a long way, though it can be supplemented by two other considerations. First, courts can consider the non-religious benefits the exemption will confer. Exemptions normally confer a benefit, but when this benefit is especially large or disproportionate to the burden imposed by the law, it may suggest the possibility that the individual might be proclaiming insincere beliefs in pursuit of those benefits. Suspicion of insincerity is particularly warranted when the purported beliefs arose only when the benefits of pursuing the exemption became known. Second, courts can look to the behavior of religious objectors to assess whether they have acted in conformity with the beliefs they assert as grounds for the exemption request. For instance, a U.S. district court once noted that a prisoner’s history of drug use and tattoos casts reasonable doubt on his purported religious objection to having his
blood drawn for DNA purposes.\textsuperscript{35} Of course, inconsistent compliance is not conclusive evidence of insincerity, as even the most sincere believers routinely fail to comply perfectly with their religious obligations.\textsuperscript{36} Still, a presumptive norm of performative consistency can aid in assessing the sincerity of a proclaimed religious belief motivating an exemption request. Third, sincerity can sometimes be assessed on the basis of group membership. Persons who are sincere members of a group that affirms particular religious beliefs can sometimes be judged, by virtue of their membership, to share in those beliefs.\textsuperscript{37} Sincerity of belief can be considered to trickle down, in a sense, from a group to its members. This should not, however, be taken to imply that sincere religious beliefs must be in conformity with the dominant beliefs in a religious community: some individuals may sincerely hold beliefs outside the mainstream of their tradition, and their disagreement with other members of the faith should not count against their sincere claim.

\textit{Burdens on Third Parties}

Any exemption necessarily entails some burden on other people. This might seem like a counterintuitive claim, since many exemption requests essentially concern persons who wish to be left alone to live as they choose. But even in these cases, there is some burden imposed on third parties – i.e. persons other than the exemptee or the legal entity granting the exemption.

One basic way in which this is the case involves the administrative resources required to facilitate an exemption. For example, in order for a person to receive conscientious objector status in regard to military selective service, some government agent must receive the prospective objector’s written request, some further agent(s) must evaluate the written request,

\textsuperscript{35} United States v. Zimmerman, 514 F.3d 851 (9th Cir. 2007).
\textsuperscript{37} Greenawalt, \textit{Free Exercise and Fairness}, 122.
and some agent(s) must carry out an in-person hearing to further assess the request. Each of
these steps requires some resources, and the cost of those resources is eventually borne by the
public whose tax contributions support the functioning of the Department of Defense. While
these administrative burdens are technically trivial – they just are what it means to deal in
exemptions – they may be materially significant. Each person-hour spent adjudicating exemption
requests is one that cannot be dedicated to other governmental functions.

Some exemptions can also create burdens that society as a whole must bear. For instance,
cases such as Smith or O Centro which grant permission for some persons to use otherwise
prohibited drugs have the potential (however minimal) to increase the incidence of drug abuse
and addiction. If this comes to pass, then communities will be required to make increased
expenditures for addiction prevention and/or treatment. Or, consider the case of Wisconsin v.
Yoder, in which the Supreme Court affirmed the right of Old Order Amish families to remove
their children from public schools after the eighth grade, in violation of compulsory education
laws. The state of Wisconsin, arguing against the exemption, raised the possibility that some
children would later choose to leave the separatist agrarian Amish community, and would then
be ill-equipped (by virtue of having missed additional years of state education) for life in society
at large. The Court rejected this argument as “highly speculative,” but this does not entirely
resolve the point: even if the risk of under-educated ex-Amish draining the state of resources is
small, it is nevertheless a risk. (If it were not, then the state would have no interest in continuing

38 Department of Defense Directive 1300.06 (May 31, 2007). Available at:
40 The Court wrote that: “This case, of course, is not one in which any harm to the physical or mental health of the
child or to the public safety, peace, order, or welfare has been demonstrated or may be properly inferred.” Ibid.,
406:230.
education beyond eighth grade for anyone.) Granting the exemption, then, means imposing this risk (however small) on the society at large.

In short, even exemptions that do not directly involve religious citizens’ conduct in relation to others do impose some burden on third-party citizens, even if it is minimal and indirect. What is significant, though, is that the burden created in these cases is widely distributed. It is distributed over an entire administrative bureaucracy, or an entire citizenry, such that the material consequences for any particular individual are miniscule. It would be practically impossible for any individual citizen to calculate their tax burden generated by the availability of conscientious objector status. From the vantage of stability, these diffuse third-party burdens are generally acceptable, insofar as they are unlikely to generate resentment in those who bear them. Every administrative program or decision imposes indirect, diffuse burdens on citizens collectively; that religious exemptions do the same does not threaten their ability to win support.

Some kinds of religious exemptions, however, will impose on third parties much more acute and appreciable burdens. Among the most obvious examples are the increasingly common cases where religious parties seek exemptions from civil rights laws protecting individuals against discrimination.41 Take, for example, the case of Thomas v. Anchorage Equal Rights Commission, where a landlord requested an exemption from a law prohibiting him from refusing to rent to a same-sex couple.42 The burden that the exemption would have imposed is obvious: any same-sex couples interested in finding housing in the area would have been required to look elsewhere than this landlord’s property. Similar burdens arise in many cases involving claims of complicity, where a religious party objects to being conscripted into cooperation with some third party’s action which they regard as sinful or prohibited (such as same-sex romantic

42 Thomas v. Anchorage Equal Rights Commission, 220 F.3d 1134 (9th Cir. 2000).
cohabitation). By exempting the religious objector from their cooperative or facilitating role, the other party is deprived of whatever benefit would ordinarily derive from the objector’s participation (such as availability of housing stock).

The material severity of this burden will vary depending on the particulars of each case. For instance, in a housing discrimination case, the burden will depend on the availability of alternative rental housing. Where the supply is ample, then the material burden may be rather light. But material burden is not the only kind of harm that a religious exemption can generate for third parties. They can also result in dignitary harms, or affronts to a person’s sense of worth or status as an equal citizen, entirely independent of the material consequences. Dignitary harms arise from laws that protect citizens in discriminating against others. Regardless of whether there are other apartments available, being denied access to a rental property on the basis of sexual orientation is offensive to a person’s dignity. In the recent case of Kim Davis, a Kentucky county clerk who refused to issue marriage licenses for same-sex couples, some suggested that couples could easily avoid any conflict by driving to a neighboring country whose clerk would issue the license. Depending on one’s precise geographic location and driving habits, it could be that going to a different county might even be more convenient than visiting

---

44 Much of the legal scholarship refers to third-party “harms,” rather than burdens. The difference may be semantic, but I think that ‘burden’ more accurately captures the salient issue. Not all burdens result in discrete harms. (E.g., the apartment-seeking couple could actually end up with a better rental situation than the one they were denied.) The point is not whether some actual harm results from the exemption, but that a third-party is subjected to some additional burden as a result of the exemption.
46 Cf. Douglas Laycock, who writes that “Hurt feelings or personal offense are so far not a basis for censorship of ideas in American law.” This, though, is a non sequitur. The point is not the offense given by the belief, but by the exceptional permission granted to act on those offensive beliefs in ways that are normally prohibited. Douglas Laycock, “Afterword,” in *Same-Sex Marriage and Religious Liberty: Emerging Conflicts*, ed. Douglas Laycock, Anthony R Picarello, and Robin Fretwell Wilson (Lanham, MD: Rowman & Littlefield, 2008), 198.
the seat of one’s home county. Nevertheless, to require that some residents travel outside their home jurisdiction to receive government services offends against the guarantees of equal citizenship.

Religious exemptions that impose acute burdens – whether material or dignitary – on third parties pose a threat to stability. Exemptions that generate only diffuse burdens are less likely to generate feelings of resentment that might arise in response to exemptions that impose acute burdens on particular persons. The difference between these two is the way in which third parties perceive the exemption. Exemptions that impose acute burdens can reasonably give rise to feelings that person A is being forced to shoulder the cost of person B’s religious beliefs. When same-sex couples are denied access to housing, they bear the burden of the landlord’s religious accommodation. This dynamic amplifies the potential of an exemption to generate resentment. Exemptions that impose acute third-party burdens should therefore stand under a stronger burden of justification than those that impose only diffuse, widely-shared burdens.

Scope of Exemption

Exemptions will vary in relation to the scope of the deviation from the rule of law they permit. Some will result in a small deviation; others will permit a larger deviation. As a general rule, the larger the deviation, the greater threat posed to political stability. We can see this most clearly by considering three dimensions in which an exemption’s scope might vary: the types of conduct which qualify as a burden on religious exercise, the size of the exempted class, and the amount of conduct permitted by the exemption.

The first dimension focuses on the grounds on which a petitioner may advance an exemption request. The second thesis advanced in section 5.1 above has to do with relieving
burdens on religious exercise. It remains, however, an open question as to what qualifies as a burden. Some are obvious: laws that prohibit some specific form of religious exercise, for example. But other burdens are less obvious. For instance, some people might claim that living in a society in which others engage in activities that they find objectionable counts as a burden. That others engage in sexual practices which I find (religiously) abhorrent may seem, to me, burdensome; life in a society in which those practices were prohibited would be (religiously) easier for me. This latter kind of objection is of markedly different kind from the straightforward case of an outlawed religious practice. To treat complaints about third-party activity as bases for an exemption would significantly expand the scope of those (purported) burdens that could ground an exemption claim.

The Supreme Court has, in multiple cases, rejected the notion that religious objections to the conduct of third-parties can warrant accommodation. In the case of *Lyng v. Northwest Indian Cemetery Protective Association* (1988), for example, American Indian groups objected to the proposed construction of a logging road through land they considered sacred. They claimed – and the U.S. Forest Service agreed – that the construction of the road would do significant damage to those sacred areas. Since this would negatively impact their religious exercise, the tribes alleged a First Amendment violation. The Court held, though, that the logging road would not substantially burden the tribes’ religious exercise, in spite of the damage it would do to the land, because it would neither coerce the Indian groups to violate their religious beliefs nor deny them an equal share in the rights afforded to other citizens.

Writing for the Court, Justice Sandra Day O’Connor noted that “government simply could not operate if it were required to satisfy every citizen's religious needs and desires.”

---

48 Ibid., 485:452.
sanctions or penalties for religious exercise are one thing, but the fact that some citizens find a
government action religiously objectionable is not sufficient to ground an exemption request.
O’Connor cited the 1986 case of Bowen v. Roy, in which the Court dismissed a religious
objection to the practice of assigning each citizen a Social Security Number. The opinion of the
Court in that case held that “The Free Exercise Clause affords an individual protection from
certain forms of governmental compulsion; it does not afford an individual a right to dictate the
conduct of the Government's internal procedures.”

It is easy to see how granting the claims of the petitioners in Lyng and Bowen v. Roy
could be politically destabilizing. First, it would be directly undermining of justificatory
impartiality, insofar as every potential law or governmental action would be subject to a religious
veto by any citizen or group thereof. Second, it would elevate the ordinary conflict of democratic
contestation – in which citizens advance competing agendas on the basis of various preferences,
including religion – to the level of settled law. Citizens who held some religious distaste for
something effected, permitted, or prohibited by a law would have grounds to invalidate that law
by means other than the democratic process. This would be severely undermining to the exercise
of collective self-governance, and would set the conditions for resentment among citizens.

The second dimension to the scope of an exemption considers the number of persons who
are likely to take advantage of a given exemption. Of course, there is no in-principle limit to the
number of persons who could take advantage of an exemption: every citizen could adopt the
religious beliefs of the original exemptee. But this is highly unlikely. Indeed, religious beliefs are
sufficiently static that the size of the exempted class can be estimated with reasonable confidence

in advance.\textsuperscript{50} Cases like Wisconsin v. Yoder or Gonzales v. O Centro, which concerned small religious sects unlikely to suddenly grow appreciably in number, could be adjudicated on the assumption that the class of persons taking advantage of the exemption would be relatively small.

When the exempted class is small, it is possible to grant an exemption while avoiding a basic concern about political order and religious exemptions: that allowing some persons to avoid the requirements of a generally-applicable law will undermine the end that the law is supposed to serve. For instance, if a democratic legislature determines (and an executive and judiciary affirm) that prohibition of possession, transportation, and consumption of some substances is a legitimate and worthy goal, then achievement of that end is undermined by an exemption permitting some individuals to possess, transport, and consume those substances. This remains the case regardless of whatever other worthy ends might be served by granting the exemption. But the degree to which the law is undermined can be minimized when the exemption is taken up by a relatively small slice of the population. The rule of law can tolerate permitting a handful of persons in a small religious community to consume hoasca tea in periodic religious ceremonies. Their numbers are so small as to barely register on the radars of most other citizens.

However, it is also possible that an exemption could cover a class so large that their dispensation from the law would be more likely to undermine the ends the law was intended to serve. A case that demonstrates the threat posed by a large exempted class is that of religious exemptions from compulsory immunization. Some states and counties in the U.S. have long

\textsuperscript{50} I leave aside here what Tushnet and others refer to as “induced beliefs,” or sincere beliefs that persons come to hold in response to the availability of an accommodation for those beliefs. Tushnet, “Accommodation of Religion Thirty Years On,” 10–11. This phenomenon seems conceptually plausible, though without evidence of its pervasiveness, I am reluctant to assume that significant changes to the size of the exempted class could occur on this basis.
provided both medical and “philosophical” or “personal belief” exemptions from laws requiring that schoolchildren be vaccinated against important diseases. Recently, however, some of those states – including California, most notably – have moved to eliminate non-medical exemptions.\(^{51}\)

One reason for the concern is that there is an objective upper limit to the percentage of persons who can be non-immunized before the phenomenon of “herd immunity” is undermined and the entire point of the vaccination regime is lost. That limit varies by disease; for measles, for instance, it is said to be less than ten percent.\(^{52}\) In some parts (principally wealthy areas) of Los Angeles County, though, 6.4 percent of the population had (as of 2014) received personal belief exemptions from mandatory immunization. This number represented a general increasing trend over recent decades.\(^{53}\) The case of immunizations presents perhaps the closest thing to an objective limit of the size of the exempted class that may be tolerated. In many cases, though, there is no fixed threshold above which an exemption poses a threat to stability. Rather, the threat increases (linearly or otherwise) along a continuum: the larger the size of the exempted class, relative to the general population, the greater the problem for stability.

What to do, then, when presented with an exemption request that, all else being equal, warrants granting, but which would likely lead to some sizeable portion of people taking advantage of it? One possible answer is simply to deny the exemption. But this solution poses a different problem: if there are so many individuals who would take advantage of the exemption were it granted, then denying the exemption leaves those persons subject to some burden on their

---


religious exercise that they cannot regard as justified. This status quo is also undermining to stability. It is a no-win situation, insofar as both granting and denying the exemption poses a threat to stability. In this case, it is worth reconsidering the burden-generating law itself. A law that generates a burden that justifies an exemption that has the potential to undermine that very same law is inherently destabilizing. The best outcome in such a case is to repeal or modify the law in order scale down the burden it imposes on religious exercise. Brian Barry advocates this approach as part of his argument against exemptions. He holds that repealing and reformulating the law will often be more net-beneficial than offering selective exemptions.54 I affirm the claim in weaker form, as applied only when the prospective exempted class is substantially large. When granting an exemption would significantly undermine the end the law is intended to serve, the better approach is to reconsider the law itself.

The third dimension along which exemptions may vary in scope concerns the range of activity exempted. This is best illustrated by an example. In 2011, the Supreme Court considered the case of *Hosanna-Tabor Evangelical v. Equal Employment Opportunity Commission*, which pitted a Lutheran school against a teacher, Cheryl Perich, who had been dismissed from the school.55 The teacher had developed narcolepsy, took a leave of absence, and several months later presented herself to return to work. At that point, she was informed that she had been replaced, and she was fired. Perich filed a claim with the Equal Employment Opportunity Commission, alleging that she was wrongfully terminated in violation of the Americans with Disabilities Act. The Supreme Court ultimately ruled unanimously in favor of Hosanna-Tabor, with their decision turning on an affirmation of the aforementioned “ministerial exception” from the Civil Rights Act. The Court judged that Perich’s duties included teaching religion (among

other subjects), which brought her under Hosanna-Tabor’s prerogative to “select its own ministers.”

In the decision, the Court declined to resolve a critical question: exactly what qualifies a person as a minister? Indeed, it avoided the issue deliberately:

“We are reluctant ... to adopt a rigid formula for deciding when an employee qualifies as a minister. It is enough for us to conclude, in this our first case involving the ministerial exception, that the exception covers Perich, given all the circumstances of her employment.”

In short, we don’t know what a minister is, but we know that Perich is one. The Court has left open the possibility of later specifying these limits, but as of this moment it has left the designation of an employee as a minister entirely to the discretion of the employer.

The decision in Hosanna-Tabor left the scope of the ministerial exception unbounded in another way. One of the lines of argument advanced by the EEOC alleged that Hosanna-Tabor used the ministerial exception as a pretext to fire her for reasons having nothing to do with religion. If the purpose of the ministerial exception is to safeguard freedom of religious exercise, the argument goes, then terminations based on reasons having nothing to do with religion would fall outside the scope of the exception. The Court rejected this argument, holding that the ministerial exception does not impose limits on the rationales behind churches’ decisions regarding selection of their ministers. In the opinion, Chief Justice Roberts wrote that this objection

56 Ibid., 565:15–16.
“misses the point of the ministerial exception. The purpose of the exception is not to safeguard a church's decision to fire a minister only when it is made for a religious reason. The exception instead ensures that the authority to select and control who will minister for the faithful ... is the church's alone.”

What these two aspects of the Court’s ruling mean is that, as a matter of settled law, there are no specific restrictions on who a church designates as a minister and on what grounds they can fire these persons. The contours of these restrictions may be filled in by later decisions, but at this point, they are limits in principle only. The scope of the exemption is, consequently, undetermined, and its significance unknown. This means that the potential for the ministerial exception to result in destabilizing deviations from the rule of law is unbounded. We can envision a hypothetical church that opportunistically designates each of its employees as a minister (a “minister” of plumbing, a “minister” of accounting, and so on), and then would enjoy unlimited latitude to fire them for any reason whatsoever.

This would undermine stability in two ways. First, it would impose a significant burden (lack of employment protections) on persons employed by religious organizations. Second, it might generate reasonable resentment among other employers who remain constrained by laws governing hiring and firing practices. This observation can be generalized to other cases: when the scope of an exemption is indeterminate, it runs the risk of permitting too much. A stable

---

57 Ibid., 565:20.

58 Justice Alito, in his concurring opinion, did propose a definition: “Any 'employee' who leads a religious organization, conducts worship services or important religious ceremonies or rituals, or serves as a messenger or teacher of its faith.” Pg. 2 of Alito concurrence (slip op). The general plausibility of this definition notwithstanding, it is hard to see how the Court could settle on a definition of ministry without wading into matters doctrinal and ecclesiological. For more on this, see the discussion of justificatory impartiality just below.
framework for exemptions will therefore aim to carefully bound the scope of exemptions, both in terms of the size of the exempted class and the range of activities covered by the exemption.

*Justificatory Impartiality*

The principal challenge posed to stability by exemptions is that they represent a deviation from the presumptive norm of distributive impartiality with respect to religious belief. But they also introduce religious considerations into the determination of legal obligations, which raises the possibility that they will involve deviations from justificatory impartiality with respect to religious belief as well. A stability-enhancing framework for religious exemptions should strive to maintain justificatory impartiality inasmuch as is possible. The notion of religious exemptions themselves is consistent with justificatory impartiality, insofar as exemptions depend on questions of fact: whether individuals do affirm religious beliefs that place them in conflict with their legal obligations. An exemption does not imply endorsement of the belief that it accommodates. But the process of adjudicating particular exemption requests is fraught with potential for justificatory partiality.

For example, Courts have on several occasions wrestled with the question of how to distinguish insubstantial burdens from substantial ones, and have recognized that making a judgment in this regard often involves taking stances on substantively religious matters. The reason is that scrutiny of a religious person’s claim that their religious practices are substantially burdened by some requirement often collapses into scrutiny of the reasonableness of their religious beliefs themselves. This dynamic arose in the Hobby Lobby case, as we will discuss.

---


below, but the classic Supreme Court examination of a burden claim was in the 1981 case of *Thomas v. Review Board*.\(^{61}\) The case concerned a Jehovah’s Witness who had been employed in plant fabricating sheet metal that was used for a variety of purposes, including military ones. When he was transferred to another part of the facility that manufactured turrets for tanks, he refused to work, on grounds that participating in the production of weapons violated his religious beliefs. Thomas quit, and the case before the Supreme Court focused on whether he was eligible for unemployment compensation benefits.

Part of the discussion in the case concerned Thomas’s formulation of his objection, specifically the distinction he drew between fabricating sheet metal, at least some of which was used for the purpose of manufacturing armaments, and manufacturing the armaments themselves. Thomas held that the former activity involved no violation of his religious commitments but that the latter did. The Indiana Supreme Court, when it ruled on the case, found these two positions inconsistent, and partly on this basis upheld the denial of unemployment benefits. The U.S. Supreme Court overturned the Indiana court’s ruling, holding that it was not the judiciary’s role to question Thomas’s estimation of the burden imposed on his religious commitments. Warren Burger, in the majority opinion, wrote: “We see, therefore, that Thomas drew a line, and it is not for us to say that the line he drew was an unreasonable one.”\(^{62}\) To do otherwise would necessarily mean affirming or denying the truth of Thomas’s religious claims – namely, that constructing turrets is religiously prohibited but fabricating sheet metal (which may be used for military purposes) is not. The Supreme Court therefore upheld Thomas’s claim, and moreover established a precedent of deference to claims of burdened religious exercise. Insofar

---


\(^{62}\) Ibid., 450:715.
as this kind of deference helps courts avoid affirming or denying religious claims, this norm is consistent with justificatory impartiality, and therefore with stability.

This does not mean, however, that there is no acceptable way to assess the substantiality of a purported burden. For one, there are two dimensions to a burden on religious practice. The first is the religious cost of secular compliance: the consequences, as estimated by the believer, of violating her religious commitments. This judgment rests inevitably on substantively religious claims. The second dimension, however, does not. It measures the secular cost of religious compliance: the penalties for violating the law in conformity with religious commitments.\(^{63}\) An assessment of the latter, then, can be achieved without violating justificatory impartiality.

Additionally, burden substantiality can be gauged by assessing the starkness of the conflict faced by the religious individual. Consider, for instance, the case of \textit{Holt v. Hobbs} (2015), in which a Muslim Arkansas prisoner requested permission to grow a beard, as his religious beliefs required and prison rules generally prohibited.\(^{64}\) For this case, the conflict faced by the prisoner was inescapable: there was literally no way for him, during his incarceration, to avoid the conflict between his religious obligations and the prison rules. In contrast, in the aforementioned case of the country clerk who refused to issue same-sex marriage licenses, the conflict arose only in virtue of the clerk’s elected office. She could have avoided the conflict by resigning her post. The law does not technically \textit{require} her to violate her religious commitments; it rather puts her to a choice between relinquishing her position and violating her religious commitments. This is not an inconsequential sacrifice, to be sure. But it serves to illustrate a difference in the severity of the burden placed on her religious exercise versus that imposed on the prisoner. Burdens that


can be avoided by reasonable sacrifices are less substantial, in one respect, than those that cannot be avoided at all.

What is important for our purposes is that these latter modes of assessing the substantiality of a burden on religious exercise are consistent with the norm of justificatory impartiality. This discussion of substantial burdens has been for the purpose of illustrating how the process of adjudicating religious exemption requests can run up against justificatory impartiality. Similar problems arise in scrutinizing other aspects of exemption requests, such as whether the religious commitment or practice that is burdened is “central” to the person’s system of belief.65 But, the above discussion also shows that there may be ways to engage in these inquiries that do not involve taking sides on substantively religious questions.

I do not mean to suggest that justificatory impartiality can be respected absolutely in every case. For instance, in order to make use of the “starkness” criterion to distinguish substantial burdens from insubstantial ones, one must rely on premises about what constitutes a “reasonable” sacrifice. Is it reasonable to expect someone to quit their job to avoid a burden on their religious belief? Answering this requires a certain degree of justificatory partiality. Or, justificatory impartiality may need to give way in order to make good on some of the other constraints I have elaborated above. For instance, to specify any limits on what kinds of employees can be categorized as ministers will likely involve adopting some premises about religious leadership and communal religious organization. So, it may not be possible to maintain justificatory impartiality perfectly in every case. But violations of justificatory impartiality should be minimized, and when necessary, should be undertaken with clear eyes as to the destabilizing potential of partiality.

Compromise

Each of the above principles discussed – preference for general scheme of exemptions, the requirement of sincerity, minimizing third-party burdens, minimizing the scope of the exemption, and maintaining justificatory impartiality – specifies some parameter within which the stability-undermining aspects of exemptions can be minimized. Having established these, it also merits noting that all exemptions have destabilizing potential, and that even in cases where they can result in a net increase in stability, the same or better outcomes might be achievable by other means. As Brian Barry argued, and I affirmed above, in some cases it is preferable to simply revise the law itself, rather than mitigate its problems via selective exemptions. That exemptions are sometimes permissible should not lead us to avoid formulating laws carefully, so as to maximize both their ability to achieve their desired ends and their ability to win endorsement from the citizens who fall under the burden of their obligations. What this requires is creative thinking about the range of possible mechanisms for achieving the desired end, in order to determine the mechanism that least restricts religious exercise. Often, there will be some compromise arrangement available that does just this.

A well-known example of creative compromise involves the City of San Francisco and a conflict with the Catholic Archdiocese of San Francisco regarding an ordinance mandating that entities contracting with the city offer benefits to employees’ same-sex domestic partners equivalent to those offered to spouses.⁶⁶ (The case occurred in the mid-1990s, well before legal recognition of same-sex marriage.) The Archdiocese objected to the ordinance, insisting that it would require Catholic employers, such as Catholic Charities, “to recognize domestic partnership as equivalent to marriage” – a recognition they were loathe to grant.⁶⁷ The

---

Archdiocese threatened to sue on religious liberty grounds, but at the same time, Archbishop William Levada reiterated his public commitment to expanding health coverage. This prompted members of the San Francisco Board of Supervisors to meet with the Archbishop and to work out a compromise. The solution they arrived at revised the ordinance to allow each employee to designate one member of their household as eligible for benefits equivalent to those offered to spouses. This arrangement made it possible for Catholic employers to comply with the law without conferring implicit recognition on same-sex partnerships – and thereby without violating their religious commitments. It might seem that the difference between securing benefits for “domestic partners” vs. “any person designated by the employee” is merely semantic, and to some degree it is. But the compromise is nevertheless significant for alleviating the subjective sense of burden felt by the Archdiocese and its members while maintaining the ability of the law to achieve its end: expanding health coverage to domestic partners.

An analogous compromise has been proposed for the county clerk who refused to issue same-sex marriage licenses. Her objection centered on the fact that marriage licenses issued by her office bore her name, which she judged to imply a personal endorsement. So, some Kentucky politicians have suggested changing the format of marriage licenses to exclude the name of the county clerk in the license-granting state. Whether this change will satisfy all parties involved remains to be seen, but it offers the potential of a compromise that will uphold the end served by the law (legal recognition of same-sex marriages) while relieving the burden that the clerk believes it imposes on her personal religious beliefs.

69 Richard Perez-Pena, “Governor-Elect Pledges to Take Clerks’ Names Off Kentucky Licenses,” New York Times, November 6, 2015. Available at: http://www.nytimes.com/2015/11/07/us/kentucky-governor-elect-vows-to-remove-clerks-names-from-marriage-licenses.html. The new Republican governor has insisted that he will do this by executive order. The previous governor (a Democrat) has maintained that the governor is not empowered to make this change unilaterally.
There is no guarantee, of course, that a compromise revision of the law will be possible in all cases. Sometimes it will not. The possibility of a compromise is greatest when the religious individual can recognize the public end served by the law as justified, even though the law burdens her exercise, or even though she doesn’t want to personally contribute to the fulfillment of that end. The San Francisco case illustrates this well: the Archdiocese affirmed the end of increasing health coverage, and so was amenable to finding some compromise arrangement. Compromise is impossible (or at least less likely) in cases when the aim of the law is flatly contradicted by the individual’s religious beliefs, and the burden generated by the law is necessary to achieving that end. In such cases, compromise is likely to be more elusive. When there simply is no compromise possible, then an exemption may be the best alternative mechanism for maintaining stability. When this is the case, then exemptions should be formulated within the parameters established by the principles set out above.

5.3 U.S. History and the RFRA Balancing Test

The real work of this chapter is normative, but the discussion will be enhanced by a brief survey of the legal lay of the land regarding religious exemptions. For the sake of simplicity, we will focus on U.S. law, though other Western liberal countries do face the challenge of religious exemptions as well. Exemptions in the U.S. are currently governed primarily by the Religious Freedom Restoration Act, or RFRA, framework. RFRA can be best understood as the product of a historical evolution on religious accommodation in U.S. law.

The first major engagement with the question of a religious exemption in U.S. law arose in the late nineteenth century, when George Reynolds, a Utah Territory resident who was secretary to Mormon leader Brigham Young, was charged under the territory’s anti-bigamy
law. Reynolds was convicted, but appealed to the Supreme Court, arguing that the First Amendment protected his exercise of his religious duty (to multiple marriages, in this case). In his original trial case, Reynolds asked the court to instruct the jury that because his multiple marriages were enacted “in pursuance of and in conformity with what he believed at the time to be a religious duty, that the verdict must be ‘not guilty.’” The Court, rejecting this argument, affirmed the legitimacy of Utah’s anti-bigamy law, and recognized that the question before them in Reynolds was “whether religious belief can be accepted as a justification of an overt act made criminal by the law of the land.” The Court ruled that the answer was ‘no’: the state may not regulate beliefs, but it is permitted to regulate actions, and the fact that someone has a religious objection to the law does not validate their defiance of it. To allow religious objectors to evade the law, the Court argued, would subordinate the civil law to both religious doctrine and individual preference. Reynolds’ conviction would therefore stand, and the doctrine was thereby established that the First Amendment provided no guarantee of an exemption for a citizen whose religious exercise is burdened by a generally-applicable law.

This doctrine held for over 80 years, until the 1963 case of Sherbert v. Verner, which concerned the denial of unemployment benefits to someone whose work obligations conflicted with her religious beliefs. In that case, a member of the Seventh-Day Adventist Church was fired from her job when she refused to work on Saturday, which she observed as the Sabbath. She then

---

70 This is not to say that cases did not arise at levels beneath the Supreme Court. For instance, Martha Nussbaum describes the 1813 case of Fr. Anthony Kohlmann, who refused to reveal the identity of a thief made known to him during the sacrament of confession. His case established the rule of confessional privilege. Martha C. Nussbaum, Liberty of Conscience: In Defense of America’s Tradition of Religious Equality (New York: Basic Books, 2008), 126–30. See also Michael W. McConnell, “The Origins and Historical Understanding of Free Exercise of Religion,” Harvard Law Review 103 (1990): 1410–12.
71 Reynolds v. United States, 98 U.S. 145 (1878).
73 Ibid., 98:162.
74 Ibid., 98:167.
applied for unemployment benefits, but was denied on grounds that she had voluntarily refused work. The Supreme Court recognized that the decision of the South Carolina Employment Security Commission imposed a burden on Sherbert’s religious exercise: it “forces her to choose between following the precepts of her religion and forfeiting benefits, on the one hand, and abandoning one of the precepts of her religion in order to accept work, on the other.”\textsuperscript{75} The Court upheld Sherbert’s claim on grounds that the state lacked any compelling interest to justify imposing this substantial burden. The case thus established the two-prong “Sherbert” test, which held that state burdens on religious exercise are only justified when they satisfy two conditions: 1) they are in furtherance of some compelling state interest, and 2) no alternative form of regulation could effectively secure that state interest without imposing an equal or greater burden on religious exercise.

The so-called Sherbert test formed the basis for religious accommodation jurisprudence for the following decades, providing a framework for courts to evaluate the legitimacy of individuals’ claims against generally-applicable laws that conflict with their religious beliefs and practices. It might seem that the ruling would have opened the door to greatly expanded grants of exemptions for religious belief, but in actuality relatively few requests for exemptions were approved by the Supreme Court.\textsuperscript{76} The most well-known of the exemptions that the Court did grant was in \textit{Yoder v. Wisconsin}, in which three Amish families petitioned were awarded an exemption from compulsory education requirements. On the whole, however, courts tended to defer to governmental claims regarding compelling state interests and did not grant an overwhelming number of exemptions.

\textsuperscript{76} Lupu, “Hobby Lobby and the Dubious Enterprise of Religious Exemptions,” 50–53.
A pivotal moment was reached in 1990, when the Supreme Court reversed its holding in *Shebert v. Verner* and returned to the position of *Reynolds*, denying that the First Amendment requires exemptions from laws that burden religious exercise. As mentioned above, the case of *Employment Division v. Smith*, concerned members of the Native American Church whose sacramental practice involved the consumption of peyote, a substance banned under the state of Oregon’s controlled substances law. Alfred Smith and his associate Galen Black were fired from their jobs for consuming peyote, and then, like Sherbert, were denied unemployment benefits because they had been discharged for “misconduct.” The Court, in an opinion written by Justice Antonin Scalia, denied Smith’s request for an exemption from the peyote prohibition, asserting that “We have never held that an individual’s religious beliefs excuse him from compliance with an otherwise valid law prohibiting conduct that the State is free to regulate.” The logic of the *Smith* decision relied on a reading of *Sherbert* that severely limited its scope. Instead of subjecting Oregon’s denial of benefits to the two-pronged *Sherbert* test, Scalia recurred to the concerns expressed over a century prior in *Reynolds*: “To make an individual's obligation to obey such a law contingent upon the law's coincidence with his religious beliefs, except where the State's interest is ‘compelling’ – permitting him, by virtue of his beliefs, ‘to become a law unto himself’ – contradicts both constitutional tradition and common sense.” The position of the Court thus reverted to a blanket denial of the claim that religious exemptions are required as a matter of protection of free exercise of religion under the First Amendment.

The Court’s denial of exemptions closed with a suggestive invitation. “Values that are protected against government interference through enshrinement in the Bill of Rights are not

---

78 Ibid., 494:885.
thereby banished from the political process,” Scalia wrote.\textsuperscript{79} The fact that exemptions are not constitutionally required does not mean that other branches of government are precluded from offering them. The Court’s decision in \textit{Smith} turned out to be widely unpopular, and only a few years later, Congress responded by taking up Scalia’s suggestion and passing in bipartisan fashion the Religious Freedom Restoration Act, which was signed into law by Bill Clinton in November of 1993. RFRA aimed explicitly at restoring the \textit{Sherbert} test, prescribing that:

(a) In general

Government shall not substantially burden a person’s exercise of religion even if the burden results from a rule of general applicability, except as provided in subsection (b) of this section.

(b) Exception

Government may substantially burden a person’s exercise of religion only if it demonstrates that application of the burden to the person—

(1) is in furtherance of a compelling governmental interest; an

(2) is the least restrictive means of furthering that compelling governmental interest.\textsuperscript{80}

\textsuperscript{79} Ibid., 494:890.
\textsuperscript{80} 42 U.S. Code § 2000bb–1.
With the passage of RFRA into law, proponents of religious liberty claimed a substantial victory, as the law established a legal presumption against burdens on religious exercise. In 1997, the court held that the law was an overreach of Congressional authority, and ruled it unconstitutional as applied to the states. Since then, many states have enacted RFRA statutes of their own. In addition to the popular backlash against Smith, legislative efforts to support religious exemptions gained support from the scholarly work of legal scholars such as Michael McConnell and Douglas Laycock, who advocated a more expansive interpretation of the free exercise clause of the First Amendment. In any event, the test enshrined by RFRA – a balancing act weighing “substantial” burdens against “compelling” interests – has been for over two decades the principal framework for adjudicating religious exemption requests in the United States, and has informed decisions regarding such practices as consumption of banned substances in religious worship, religiously-motivated killing of bald eagles, and refusal to provide insurance for contraceptive drugs to one’s employees. It is to this last case that we now turn.

5.4 The Case of Hobby Lobby

One of the most controversial and anxiously-awaited cases in the Supreme Court’s 2014 term was the case of Burwell v. Hobby Lobby. The case was one in a series of cases challenging the legality of President Obama’s signature legislation, the Affordable Care Act (ACA)

---

84 United States v. Aguilar, 527 F. App’x 808, 812-13 (10th Cir. 2013).
healthcare reform law.\textsuperscript{85} Specifically, the case brought by Hobby Lobby, which is a chain of craft stores operating mostly in the American South, challenged the ACA on religious liberty grounds, alleging that the law infringed on the religious exercise of the Green family, owners of the Hobby Lobby chain.\textsuperscript{86} Among the ACA’s provisions is a requirement that employer-provided healthcare meet certain standards of “minimum essential coverage,” which includes (among much else) any methods of contraception approved by the Food and Drug Administration. The Green family claimed that four of those methods work by preventing the implantation of a fertilized egg, and consequently amount to “abortifacients,” or abortion-causing drugs. The Greens are conservative Christians who oppose abortion, and they claimed that the minimum coverage requirement would force them to violate their religious commitments by facilitating their employees’ abortions by means of providing insurance coverage for abortifacient drugs. The consequence for failing to do so would be a penalty of $100 per day per each affected individual, or around $475 million per year.

The Court ruled 5-4 in favor of Hobby Lobby, asserting that the government had an obligation to provide them with an exemption from the contraception coverage requirement. The framework under which the ruling was reached was that of the Religious Freedom Restoration Act of 1993 (RFRA), which we will consider in detail in the following section. For now, though, we will set aside the RFRA-specific particulars of the case in order to make a few observations about the decision. The first thing to note is that, contrary to the allegations of some of the decision’s critics, the Court did recognize (albeit \textit{arguendo}) the provision of contraception coverage as a compelling public interest.\textsuperscript{87} That is, its grant of an exemption was not based on a


\textsuperscript{86} The following case facts come from \textit{Burwell v. Hobby Lobby}.

\textsuperscript{87} Burwell v. Hobby Lobby Stores, 573 U.S. ___, 40 (2014), 573.
judgment that access to contraception is unimportant or unworthy of enshrinement in public law. Second, the crux of the decision was the following observation: the government had already devised an exemption for non-profit organizations with religious objections, and there was no reason why the same accommodation could not be extended to Hobby Lobby. Justice Alito, in the majority opinion, wrote: “HHS [the Department of Health and Human Services] itself has demonstrated that it has at its disposal an approach that is less restrictive than requiring employers to fund contraceptive methods that violate their religious beliefs.” This alternative arrangement “does not impinge on the plaintiff's religious belief that providing insurance coverage for the contraceptives at issue here violates their religion, and it serves HHS's stated interests equally well.”

Thus, Hobby Lobby was granted the exemption that already was offered to some other religious objectors to the ACA’s contraception mandate.

There is a great deal that might be (and has been) said about the Hobby Lobby case. What matters for our purposes is how the decision stands in relation to the end of stability. Is the exemption granted in *Hobby Lobby* likely to be net supportive of stability, or net undermining to it? I won’t attempt a definitive answer to the question of whether *Hobby Lobby* was rightly decided, but we can gain purchase on its implications for stability by evaluating the ruling in light of the parameters set out in the above section.

First of all, while the Court recognized that the contraception mandate is justified by compelling public reasons (public health and gender equality), it imposes a burden that the owners of Hobby Lobby are unlikely to be able to regard as justified. The Green family believes that “human beings deserve protection from the moment of conception, and that providing

---

88 Ibid., 573:43–44.
insurance coverage for items that risk killing an embryo makes them complicit in abortion.”" It is therefore unlikely that, from within the perspective of their conception of the good, they will be able to regard as justified a law obligating them to participate in their employees’ use of abortifacient drugs. The case then falls into the category of cases where an exemption may be warranted. Additionally, and uncontroversially, there is no obvious reason to doubt the sincerity of the beliefs claimed by the Green family, owners of the Hobby Lobby chain. Neither the government nor the Justices raised any question about sincerity, nor were there any apparent grounds for doing so. Indeed, Hobby Lobby has demonstrated its commitment to conducting business within the parameters of its beliefs in other ways, such as refusing to sell shot glasses in order to avoid promoting alcohol use. It is fully appropriate, then, to regard Hobby Lobby’s claim of a conflict between religious and legal obligations as sincere and legitimate.

The *Hobby Lobby* exemption also meets the criterion of minimizing third-party burden. In fact, extending to Hobby Lobby the exemption already provided for nonprofit organizations does not impose *any* acute burden on third parties. The relevant parties, of course, are those female employees of Hobby Lobby who might take advantage of contraception coverage. In a dissenting opinion, Justice Ruth Bader Ginsburg raised the concern that “The exemption sought by Hobby Lobby ... would deny legions of women who do not hold their employers' beliefs access to contraceptive coverage that the ACA would otherwise secure.” But this worry was ill-founded: the accommodation devised for nonprofits included an alternative mechanism for guaranteeing contraceptive coverage. In the case that an employer objects to paying for coverage for the drugs, the insurer itself must provide the coverage, without imposing any cost-sharing

---

91 See Justice Ginsburg’s dissent, at pg. 8.
requirements on the employer or the covered employees.\textsuperscript{92} Thus, the exemption does not threaten anyone’s access to contraceptive coverage, and thus imposes no burdens on third parties.\textsuperscript{93}

Assessing the scope of the exemption is less straightforward. In one respect, it is a relatively narrow exemption, applying only to payment arrangements for insurance coverage of a particular healthcare good – birth control. On the other hand, the class of entities that could take advantage of the exemption is harder to identify. The decision limited itself to application to “closely-held” corporations, which are defined by the IRS as corporations more than half of whose stock is owned by five or fewer individuals.\textsuperscript{94} Ninety percent of for-profit corporations in the United States are closely-held, though because many of these are very small, only fifty-two percent of American workers are employed by closely-held corporations. That is a large portion of the workforce. But, on the other hand, exemptions such as that granted to Hobby Lobby would be limited to those corporations that have a sincere religious objection to the law. (For now, set aside the question of whether corporations can have religious objections; we will return to it below.) Depending on how strict a standard for sincerity one adopts, this could limit the scope of the exemption quite significantly.\textsuperscript{95} On the other hand, yet again, some have questioned whether the Court’s logic can be restricted to closely-held corporations, or if it must eventually be

\textsuperscript{92} Burwell v. Hobby Lobby Stores, 573 U.S. __, 43 (2014), 573.
\textsuperscript{93} As Alito wrote, “The effect of the HHS-created accommodation on the women employed by Hobby Lobby and the other companies involved in these cases would be precisely zero.” Ibid., 573:4. Also Amy J. Sepinwall, “Conscience and Complicity: Assessing Pleas for Religious Exemptions After Hobby Lobby,” University of Chicago Law Review 82 (2015): 184. Cf. Andrew Koppelman and Frederick Mark Gedicks, “Is Hobby Lobby Worse for Religious Liberty Than Smith?,” St. Thomas Journal of Law and Public Policy, forthcoming. Koppelman & Gedicks’s argument relies on the assumption that the Court would have granted the exemption even had the alternative accommodation mechanism not been in place. This seems highly speculative, and stretches the language of the decision to its most extreme interpretation. NeJaime & Siegel raise concern that the accommodation might give rise to dignitary harms, but I’m not entirely convinced by this either. See “Conscience Wars: Complicity-Based Conscience Claims in Religion and Politics,” 2581–83.
\textsuperscript{95} Since Hobby Lobby had previously demonstrated its religious commitments (e.g. in avoiding promotion of alcohol), the case could set a precedent for a high bar for corporate religious claims, requiring something along the lines of the “performative consistency” standard I proposed above.
extended to all for-profit entities. This move would of course expand the number of corporations that could request exemptions. But the sincerity requirement would still stand, and it would likely be more difficult for a corporation owned by a large number of people (such as a publicly-traded corporation) to demonstrate religious commitments.

The broader question regards the scope of the presumption in favor of exemptions that the decision established. Some have suggested that the *Hobby Lobby* precedent would seem to ground a claim to a corporate exemption from any kind of law whatsoever. Insofar as this is the case, though, it is a feature not of this particular case, but rather of the RFRA exemption framework on which the case is premised. The RFRA framework establishes a general (rather than targeted) basis for exemption claims, and the open-endedness of that framework is part of the riskiness of a general approach to exemptions. Moreover, the RFRA framework does not guarantee exemptions, but rather grants legitimacy to exemption requests. As Justice Kennedy noted in a concurrence to the *Hobby Lobby* decision, nothing in the decision implies that religious exercise should be permitted to “unduly restrict other persons, such as employees, in protecting their own interests, interests the law deems compelling.” In other words, the scope of the exemption is bounded by the compelling interests of the state and the interests of third parties.

Finally, we can note that the *Hobby Lobby* decision maintains justificatory impartiality by declining to make any judgment on the validity of the substantively religious claims put forward by *Hobby Lobby*. As we noted above, one place this can arise is in the assessment of the substantiality of a purported burden on religious exercise. In this case, *Hobby Lobby* claimed that a requirement to pay for insurance coverage that might pay for an employee’s contraception

---

98 *Burwell v. Hobby Lobby*, Kennedy concurrence, pg. 4.
that might cause an abortion constituted an infringement on their religious beliefs. Some, including the four justices who signed Justice Ginsburg’s dissenting opinion, cast doubt on the tenability of claim. She wrote that “the connection between the [plaintiff’s] religious objections and the contraceptive coverage requirement is too attenuated to rank as substantial.”

Justice Alito, author of the majority opinion, responded to Ginsburg’s assessment by noting that her estimation of the religious burden imposed by the law rested on a judgment about the reasonableness of Hobby Lobby’s belief that providing coverage for contraception would make them complicit in abortion. There is no way to make such a judgment without also judging the truth value of the petitioners’ religious claims – i.e. violating justificatory impartiality. The majority opinion, in contrast, establishes the substantiality of the burden on religious exercise by deferring to Hobby Lobby’s characterization of their religious beliefs and by noting the severe secular (financial) consequences they would face for acting fully in accordance with those beliefs. In taking this tack, the decision remains impartial with respect to the truth or falsity of the substantive religious claims advanced by Hobby Lobby.

On balance, it would seem that Hobby Lobby’s case was a good candidate for an exemption. It was based on a sincere religious belief that Hobby Lobby judged to conflict with a legal obligation. A similar exemption was already available to other parties with similar religious objections. The potential for harms to third-parties was minimal or null. The scope of the exemption, while not perfectly bounded, is limited in ways that reduce the likelihood that so many corporate employers will take advantage of it that the purpose of the law will be undermined. And while it is tempting to question Hobby Lobby’s claim that providing insurance coverage burdened their religious exercise, justificatory impartiality requires deference to Hobby

---

99 Burwell v. Hobby Lobby, Ginsburg dissent, pg. 23.
Lobby’s judgment in this respect. Each of these considerations thus leans in favor of granting the exemption, as the Supreme Court did.

We cannot conclude a discussion of *Hobby Lobby*, though, without touching on the aspect of the case that drew the most attention and controversy: the corporate status of the “person” claiming a religious exemption. Or, more specifically, Hobby Lobby’s *for-profit* corporate status. The Supreme Court had not previously extended rights of religious accommodation to *for-profit* entities, and the prospect of doing so raised a good deal of popular opposition and controversy, particularly in the wake of the 2010 decision in *Citizens United v. FEC*,100 which affirmed independent corporate political expenditures and was widely perceived as an expansion of corporate status and rights. Granting Hobby Lobby’s religious exemption requests, it was suggested, would represent a further step in the direction of corporate dominance of the American social and political landscape.101 As we discussed above, the Court’s decision took a middle path, granting that closely-held corporations could express religious claims without granting the same to any and all corporations.

From the vantage of stability, the question of whether corporations can claim religious beliefs or rights of religious exercise in a manner equivalent to human persons is something of a red herring.102 What matters for stability – that is, for the likelihood of a political order winning the endorsement of the persons living under it – is the burden a law places on the religious

---

101 See, e.g. Julia Mirabella and Sandhya Bathija, “*Hobby Lobby v. Sebelius*: Crafting a Dangerous Precedent,” Center for American Progress: “If the Supreme Court were to side with Hobby Lobby, it would be another piece of pro-corporate precedent from an increasingly pro-business court.” Available at https://www.americanprogress.org/issues/civil-liberties/report/2013/10/01/76033/hobby-lobby-v-sebelius-crafting-a-dangerous-precedent/.
exercise of the *owners* of the for-profit business. The majority opinion recognized that Hobby Lobby’s corporate claims are only proxy claims for the people who own it, writing that

> “the purpose of [the corporate structure] is to provide protection for human beings. A corporation is simply a form of organization used by human beings to achieve desired ends. … When rights, whether constitutional or statutory, are extended to corporations, the purpose is to protect the rights of these people. … And protecting the free-exercise rights of corporations like Hobby Lobby … protects the religious liberty of the humans who own and control those companies.”[103]

Does the law put the Green family – owners of the corporation – to a choice between their religious and legal obligations? In the business or commercial realm, this is a somewhat trickier question, since nobody is *forced* to own and operate a business like Hobby Lobby. But, as we discussed above, addressing the Green family’s conflict by telling them to abandon their business is not likely to resolve their feeling of being burdened by the law. Insofar as they can reasonably expect to engage in business while maintaining their religious beliefs, the ACA imposes a burden on the Greens’ religious exercise. Thus, accommodating Hobby Lobby’s religious claims is really a matter of accommodating the claims of Hobby Lobby’s owners. As far as stability is concerned, then, that Hobby Lobby is a for-profit corporation is irrelevant to the legitimacy of their religious claims.

There is one additional consideration, however, that we have not touched on so far in this discussion, and which would lean against an exemption for Hobby Lobby. That is the social and political context in which this particular exemption request arose, and the social meaning of

---

Hobby Lobby’s pursuing their case. In fighting for an exemption from the Obamacare contraceptive mandate, Hobby Lobby came to represent a principal front in two broader socio-political movements: 1) conservative opposition to the Affordable Care Act, and to President Obama himself; and 2) conservative resistance to the gradual social normalization of casual sex. Douglas NeJaime and Reva Siegel argue that exemption requests that are, like Hobby Lobby’s, based on claims of complicity in religiously-objectionable behavior are often “entangled in long-running ‘culture war’ conflicts that break with traditional morality.”¹⁰⁴ In some cases, pursuit of religious exemptions is not aimed so much (or at least exclusively) at securing the conditions for individual religious fidelity, but rather at advancing a social and political agenda. “In seeking an exemption,” they write, “a claimant need not withdraw but instead can employ the religious objection to criticize norms governing the entire community.”¹⁰⁵ The exemption can thus become not just an end in itself, but also an avenue for continuing a political fight.

This dynamic has consequences for stability. The goal of exemptions as defended in this chapter has been to ameliorate citizens’ subjective sense of burden. Exemptions, when granted strategically, can enhance stability by removing obstacles to citizens’ endorsement of the regime. Exemption requests that are pursued as tactical weapons in the culture wars run the risk of undermining stability, rather than supporting it. As NeJaime and Siegel note, these kinds of exemption requests can serve to “extend, rather than settle, conflict about social norms in democratic contest.”¹⁰⁶ This phenomenon can be observed especially in the case of same-sex marriage in the U.S.: as an increasing number of states, and eventually the federal government, came to recognize and guarantee access to same-sex marriage, conservative opponents of expanding marriage rights have turned their attention to the ramifications of these changes for

---

¹⁰⁴ NeJaime and Siegel, “Conscience Wars: Complicity-Based Conscience Claims in Religion and Politics,” 2542.
¹⁰⁵ Ibid., 2552.
¹⁰⁶ Ibid., 2520.
religious liberty.  

Or, NeJaime and Siegel show how healthcare refusal laws have served as a mechanism for advancing an anti-abortion political agenda by reducing access to abortion services. What this suggests is that exemptions such as these may not succeed in serving to bring dissenting citizens on board with a political order that includes laws they disagree with, but rather may reinforce their disaffection from it.

Additionally, exemptions of this sort are prone to generate resentment in others. First, because other citizens may also disagree with some laws, but swallow their dissatisfaction and resolve to comply, rather than seeking special accommodation. Second, citizens who inhabit other wings of the political spectrum may resent their opponents pursuing political ends in disingenuous ways, posing as embattled victims of religious oppression when they are, in reality, active and determined culture warriors. The fact that many of the parties most loudly demanding accommodations in recent years have been Christians – members of a demographically majority and culturally dominant religious group – supports some suspicion that their principal concern is not actually that “religious liberty is under attack,” but rather than they and their allies are losing ground in ordinary politics, and the rhetoric of religious liberty is a strategy to reclaim it.

Why, though, should there be any problem with efforts to advance a political agenda using whatever means possible, including the mechanism of religious exemptions? Isn’t this simply part and parcel of robust democratic engagement? On one hand, this is basically correct: creativity, and even disingenuousness, in politics is not a priori wrong. On the other hand, the framework for exemptions that I have defended here occupies a tenuous space within liberal

108 NeJaime and Siegel, “Conscience Wars: Complicity-Based Conscience Claims in Religion and Politics,” 2554–47.
politics, both enhancing and undermining stability in certain ways, and depends upon the maintenance of a certain balance in order to sustain its net contribution to stability. When parties use exemptions as a way to advance a political agenda, it threatens to undermine popular support for the idea of religious accommodation in general, and detracts from the ability of an exemption framework to support stability. An insight along these lines has been raised by prominent scholar and defender of religious liberty Douglas Laycock, who has suggested that a too-broad interpretation of religious accommodation laws will prompt legislators to withdraw their support for such laws, leading ultimately to less protection of religious exercise.111

When people exploit the availability of exemptions to advance a political or social agenda, it threatens to diminish the stability of the exemption regime itself. The fact that Hobby Lobby’s exemption request had the appearance of being politically strategic in this way would lean against granting the exemption. Like the other criteria discussed above, this feature is not independently conclusive; it alone does not necessarily demand the rejection of Hobby Lobby’s request. That said, it should weigh fairly heavily, since the element of political strategy calls into question the potential of the exemption to accomplish the very end it is intended to: enhancing stability.

The piece of evidence that most clearly suggests that the Hobby Lobby exemption did not, in fact, contribute to stability amid disagreement over contraception and sexual mores comes in the dozens of successor cases to Hobby Lobby in which religious entities allege that even the accommodation granted to nonprofits (where contraceptive coverage is provided by the insurer with no cost-sharing from the employer) imposes an unsupportable burden on religious exercise.

While Hobby Lobby appears to be contented by the ruling in their case, there are other religious entities occupying a similar place on the political spectrum who are not satisfied. Only a few days after handing down the *Hobby Lobby* ruling, the Supreme Court issued an emergency injunction blocking the IRS from imposing fines on Wheaton College, an evangelical Christian school in Illinois, which objected to the alternative contraceptive coverage arrangement. The school claimed that “its filing of a self-certification form [to communicate its contraception objection to its insurance provider] will make it complicit in the provision of contraceptives by triggering the obligation for someone else to provide the services to which it objects.” Since that time, thirty-seven distinct petitioners – including the Little Sisters of the Poor, the Roman Catholic Archbishop of Washington, Oklahoma Wesleyan University, and an organization called Reaching Souls International, Inc. – have filed similar claims. Though the details of each petition vary slightly depending on the nature of the organization and the structure of its health insurance coverage, what they share in common is an allegation that even the accommodation provided – and affirmed as a resolution to Hobby Lobby’s objection – imposes a substantial burden on religious exercise.

In these cases, it seems all the more obvious that the plaintiffs are leveraging the exemption mechanism in order to extend their struggle against sexual norms that they disdain and a law that grants those norms some implicit (though minimal) recognition. As the Government sensibly notes in one of its briefs to the Supreme Court on this case, there is no plausible sense in which the actions required of the objecting employers (which amount to “the written equivalent of raising a hand,” in the words of one appeals court ruling) force the

---

112 The Greens were quoted as saying “Our family is overjoyed by the Supreme Court’s decision,” which they called a “victory, not just for our family, but for all who seek to live out their faith.” Quoted in a Becket Fund press release titled “Supreme Court Victory for Hobby Lobby and Religious Freedom.” Available at “http://www.becketfund.org/hobbylobbyvscotusvictory/.

employer to become complicit in the independent actions of their employers.\textsuperscript{114} Indeed, the logic of the petitioners’ claims would suggest that they would maintain a religious objection to any system whatsoever that guaranteed contraceptive coverage for their employees against their objection. This suggests that what the petitioners are objecting to is not their alleged complicity in their employees’ use of birth control, but rather to the government’s policy of guaranteeing cost-free access to contraception and the subsequent use of contraception that policy is likely to facilitate.\textsuperscript{115} This conclusion, in turn, suggests a sort of insincerity in the petitioners’ claims: not that their religious beliefs are insincere, but that their stated desire (to avoid a substantial burden on their religious exercise) is not their true goal. Moreover, it is likely that overcoming these petitioners’ objections would be difficult without imposing acute burdens on third parties (i.e. women whose access to paid contraceptives would be lost or hindered), and therefore undermining the very purpose of the law. Finally, the government has argued that the petitioners’ claims are closer to the claims in \textit{Lyng} and \textit{Bowen v. Roy} than those in \textit{Smith} or \textit{O Centro}: they involve an objection to the conduct of third-parties – the government and the insurance companies – rather than to any cognizable burden imposed on the petitioners themselves.\textsuperscript{116}

Thus, the successor cases to Hobby Lobby fail a number of the tests for exemptions that are likely to serve stability. But perhaps most basic is their blatant aim to contribute to social conflict over sexual morality and the role of government. The phenomenon of politically


\textsuperscript{115} \textit{Priests for Life v. HHS}, Government brief, pgs. 17-20.

\textsuperscript{116} One thing worth noting about this claim is that it does not rest on any evaluation of the petitioners’ religious claim of a substantial burden. Just as the government conceded in \textit{Lyng} that the tribes’ religious exercise would be negatively impacted, the government can concede that the employers whose workers will receive contraception can reasonably maintain religious objections. But these objections are not ones that can be leveraged as grounds for accommodation in a stable political system. This way of framing the issue avoids the problem regarding justificatory impartiality raised by the \textit{Thomas} ruling.
strategic exemptions therefore both explains why the *Hobby Lobby* decision has not quieted conflict over the federal government’s mandate for contraception coverage, and suggests that the exemptions requested in successor cases should not be granted.

5.5 A New Test: Comparing the Stability and RFRA Frameworks

The first two sections of this chapter extended to the issue of religious exemptions the account of liberal impartiality that we have developed on the basis of stability. The third section introduced the legal framework, defined by the Religious Freedom Restoration Act, that currently governs religious exemptions in the U.S. context. The fourth section considered a recent high-profile case decided by the Supreme Court on the terms supplied by RFRA, and evaluated that decision by the standards of the normative stability-based framework. Having undertaken that investigation of a particular case, our next step is to identify precisely how the stability approach affirms, deviates from, or contributes to the RFRA framework.

Recall that the RFRA framework indicates three conditions for granting a religious exemption. First, (1) the law from which an exemption is requested must impose a *substantial burden* on the religious exercise of some person. This triggers a presumption in favor of an exemption, unless the burden-generating law can be shown to (2) serve some *compelling government interest*, and (3) be the *least restrictive means* available for pursuing that interest. If a law imposes a burden while satisfying these latter two conditions, then no exemption is warranted. If, on the other hand, a burden-generating law fails the second and third conditions, then an exemption must be granted.

The substantial burden criterion is captured by the second thesis I advanced in regard to exemptions in section 5.2. There, it was clear that the only reason to even contemplate
exemptions is in the case of substantial burdens on religious exercise. It is this burden that threatens to undermine individuals’ endorsement of the regime. Like the RFRA framework, I didn’t offer a specific or definite standard for what constitutes a substantial burden. But the basic idea is captured: unsubstantial burdens on exercise are part and parcel of life in civil society, but burdens of a substantial degree may give rise to a legitimate exemption claim. The second thesis, though, invoked an additional criterion: unjustified burdens. Two ways in which a burden could be unjustified include by a) not being supported by a compelling public reason, and b) not employing the least burdensome means to achieve its intended aim. These, of course, track closely on the second and third criteria of the RFRA standard. I did, though, also note the possibility of burdens that a person could not regard as justified simply because of a basic incompatibility between their religious beliefs and the end served by the law. In this respect the stability framework goes somewhat beyond RFRA in admitting the possibility that a narrowly-tailored law serving a compelling public end – i.e., one satisfying RFRA criteria (2) and (3) – might still fail to win endorsement, and could warrant an exemption.

It is also worth noting that the RFRA framework accords with two of the further principles offered in 5.3. First, it provides a general framework for exemptions, establishing grounds for any person to request an exemption from any legal requirement on the basis of any religious belief, rather than trying to identify or “target” specific practices or beliefs for exemptions. Second, the RFRA framework assumes the sincerity of the religious beliefs in question. This is not only part of the Court’s process of examination, but it is also implied in the notion of a substantial burden: a law that burdens a religious belief that someone claims but does not in truth hold does not actually impose a substantial burden on their religious exercise. In short, the basic RFRA framework largely accords with the stability approach.
The major difference, though, is that the RFRA framework stops short where the stability approach supplies additional normative principles for discriminating among exemption requests.

First, exemptions should be granted only in accordance with a principle of accommodative equity. Because it speaks of religion in general terms, RFRA implies equity among different religions. Whether religion may, or must, be granted special solicitude over against other considerations is of greater dispute. Some scholars argue that the constitution requires strict neutrality between religion and non-religion, while others claim that courts and legislatures are free to prefer religion.117

Second, exemptions that impose acute third-party burdens should be minimized. Consideration of third-party harms is not explicitly a part of the RFRA framework, and there is disagreement among scholars over the degree to which precedent on religious accommodation calls for minimizing harms to third parties.118 From the vantage of stability, though, this principle is critical, for it avoids creating new burdens that offset the gains made by granting the exemption in the first place.

Third, exemptions whose scope is either large or indefinite with a high probability of being large should be minimized. RFRA does not provide any explicit mechanism for limiting the scope of exemptions, though this criterion might be understood to fall out of the compelling interest requirement. This approach confronts the issue only indirectly, though.


Fourth, exemptions must be adjudicated on grounds that maintain justificatory impartiality. Some legal scholars and jurists have argued that judicial neutrality with respect to religious claims is required by the Establishment clause of the First Amendment, but the disagreement between Justices Alito and Ginsburg makes clear that there is no consensus on what kind of scrutiny of religious claims is permissible by judges. More to the point, the norm of justificatory impartiality is not codified anywhere, and it is certainly not indicated specifically by RFRA.

Fifth, stability warns against granting exemptions that are part of political agendas designed to extend social and political conflict beyond the limits of ordinary democratic politics. RFRA obviously does not include anything like this provision. Admittedly, this is a hard principle to judicially enforce, because determining whether an exemption is being pursued for its own sake or for the sake of some broader agenda (or both) will always be a matter of contingency and judgment. But it is nevertheless a normative principle that supplies a general guideline, however difficult to actualize.

Sixth, and finally, the stability approach relegates the five foregoing principles to a secondary status, prioritizing compromise solutions when possible. RFRA, of course, does not rule these out, and parties in any given exemption context might have their own reasons for preferring a compromise to an exemption. But neither does the RFRA framework supply the normative preference for compromise that the stability approach entails.

In sum, the stability approach represents an improvement over the RFRA framework for religious exemptions in two general ways. First, it admits of the possibility of exemptions even when the burden-generating law is narrowly tailored to advance a compelling public interest. Second, it provides more substantive and specific criteria for discriminating among exemption

---

requests – that is, for approving some and rejecting others – than the RFRA balancing test does. Recall that in the Court’s opinion, written by Justice Scalia, in *Employment Division v. Smith*, the issue was that the question of religious exemptions seemed to be all-or-nothing: either exemptions must be denied categorically or there can be no way of discriminating among requests of all kinds. The Court chose the former path, which prompted the pendulum to swing the other way, with RFRA supplying a presumptive justification for exemptions in any case of substantial religious burden. What the stability approach provides is a middle path between ‘all’ and ‘nothing,’ supplying parameters for granting exemptions that are ultimately enhancing to political stability, and rejecting those that are not.

### 5.6 Some Challenges

In concluding this examination of religious exemptions and my defense of a stability-based normative framework for dealing with them, I want to consider four questions and potential objections to the argument I’ve advanced here.

*Is Religion Special?: Conscience*

The argument of this chapter has worked to establish the parameters within which religious exemptions can be supportive of stability. The further question naturally arises: what about exemptions for beliefs and commitments other than religious ones? Does the goal of stability justify giving special deference to religion, or should we actually be talking about a broader category of “conscience” exemptions?\(^{120}\) In the U.S. context, religion is constitutionally privileged, in the sense that the First Amendment explicitly protects “the free exercise of

---

religion,” and does not mandate similar protection for non-religious ethical beliefs or commitments of conscience. A number of legal scholars have disputed that religion ought to be “singled out” for special treatment – one has even suggested that the First Amendment may be “morally defective” for falsely assuming the specialness of religion. Our concern here, though, is not with the status assigned to religion by virtue of its mention in the text of the Constitution, but rather its significance vis-à-vis stability.

To answer the question of whether the stability approach would also sanction exemptions for non-religious commitments of conscience, we have to identify what it is about religion that matters for the stability approach. There are two features of religion that matter for the argument we’ve advanced here. First, religious commitments are highly valued: people sufficiently value their religious commitments that they are normally likely to regard infringements on those commitments as a significant bad, one that is likely to undermine their endorsement of the regime. Second, religious commitments are strongly held: persons hold their religious commitments sufficiently strongly that when a person is able to regard a regime as justified from within the perspective of their religious beliefs, they are normally likely to sustain their support for the regime against temptations to noncomply.

The question is, then: do non-religious beliefs and commitments of conscience satisfy these two conditions in such a manner as to render them deserving of accommodation? The answer is actually quite straightforward: yes. Or, more precisely, there are some commitments of conscience that are both highly valued and strongly held. Laws that burden them are likely to

---

undermine citizens’ support for the regime, and these commitments are capable of motivating support just as religious commitments can. Given this, the stability approach would justify conscience exemptions on the same grounds that it justifies religious exemptions.

The only problem that arises in granting exemptions for non-religious beliefs is that it is difficult to determine exactly which beliefs fall under the umbrella of conscience and which do not. For instance, is my commitment to recycling a matter of ‘conscience,’ or just a preference, albeit a value-laden one? Identifying and enforcing the line between conscience and non-conscience is difficult, and the broader the category of conscience is drawn, the harder it becomes. Though, this boundary problem is, at least conceptually, no more or less vexing than determining the bounds of what does and does not count as ‘religion.’ These are serious challenges to a coherent regime of accommodation for religion and/or conscience, but from the vantage of stability, either category is in actuality just a proxy for the class of commitments that meet the conditions of being highly valued and strongly held. Religious commitments are a paradigmatic example of this kind of commitment, but in truth any commitment that meets the two conditions would, by the terms of the argument offered here, warrant exemptions from burden-imposing laws.

Is Religion Special?: Association

Though it hasn’t played a prominent role in this chapter, one of the reasons that the stability account identifies in support of religious exemptions focused on the role that moral communities, such as religious associations, play in cultivating the kind of deep commitments that can motivate reliable support of a regime. A regime of liberal impartiality may, however

---

inadvertently, impose burdens on the maintenance of distinctive moral communities. Exemptions to relieve these burdens can be indirectly supportive of stability by permitting those communities to flourish. Along the same lines as the question of whether the stability argument for exemptions could apply to non-religious beliefs, we may also wonder whether the stability approach could ground presumptions in favor of exemptions for non-religious associations.

The Supreme Court has considered the extent of the latitude that ought to be afforded to voluntary associations, most prominently in the case of *Roberts v. United States Jaycees* (1984).124 The “Jaycees,” or the national Junior Chamber of Commerce, was a business-oriented civic association that, prior to the Supreme Court decision, restricted full membership to young men. The question of membership, and latitude to discriminate in manners that are ordinarily prohibited, is one of the basic issues that arises in determining the proper scope of associational liberty. The Court ruled unanimously against the Jaycees, holding that the Jaycees were neither an “intimate” nor “expressive” organization – the former referring to a small, private association constituted by personal relationships; the latter describing associations formed for the purpose of expression or speech – and therefore did not merit protection from the government’s anti-discrimination regulation. This was the constitutional resolution; what would the stability approach have to say about the Jaycees case?

We should recall that in the Hobbesian argument in chapter 3, we identified associations as a key threat to stability. So, measures to encourage and support associations must be balanced against their potentially destabilizing role. Then, in order to determine whether the end of stability would justify exemptions for non-religious associations, we must once again identify the salient features of religious associations and ask whether they apply to non-religious associations as well. There are two such features. First, religious associations effectively cultivate deep (that

---

is, highly valued and strongly held) commitments. They accomplish this both by their formative and educative functions and by providing a forum in which individuals can have their commitments nourished and sustained. Second, there are some aspects of religious exercise that can occur only in association with others, such that some persons will perceive burdens on collective religious activity as burdens on their individual religious exercise. Group associations may therefore be warranted to alleviate those individual burdens. What this means is that, from the vantage of stability, religious associations are valuable not by virtue of their “expressive” dimensions, nor any value “intrinsic” to association, as authors such as George Kateb have argued.125 Rather, it is their instrumental role developing individuals’ religious commitments and supporting their exercise that makes them valuable, and which may warrant some exemptions.

On the basis of these two features, are there analogous non-religious associations that would be similarly deserving of accommodation? On the educative and formative function of religious communities, my suspicion is that there are not many other associations that can reliably cultivate deep commitments in the way that religious communities can. This is not to say that there are no other organizations that perform educative functions. George Kateb is of course right about this: “Parties, advocacy groups, business enterprises, and private educational institutions … specialize in ‘cultivating and transmitting shared deals and beliefs’…”126 But there are few other kinds of organizations that cultivate and transmit beliefs as reliably as religious organizations. Political parties, for instance, of course have a platform that they aim to communicate to members (and others), but those platforms are constructed in part with an eye to attract voters as they are, rather than to form them. Indeed, part of membership in a political party is contributing to deliberation over what the platform ought to be. Thus, while parties may

---

126 Ibid., 47.
transmit beliefs, it is equally the case that they instantiate the beliefs that citizens already hold. Moreover, most voluntary associations do not engage their members’ beliefs and commitments at the level of depth that religious communities do. It is hard to imagine anyone suggesting that the Jaycees – or even an ethos-driven organization like the Boy Scouts\textsuperscript{127} – can cultivate commitments with a reliability comparable to religious communities’ cultivation of commitments. If the justification for exemptions on the basis of religious communities’ educative capacities can be translated to non-religious associations, it is only for the rare association that reliably exerts a formative effect at the level of depth that religious communities do.

The second important feature of religious associations is more readily translatable. Religious associations, in some cases, provide necessary conditions for the realization of a basic good – namely the exercise of religious beliefs. The subsequent question is: what other associations provide necessary conditions for the satisfaction of some good of such basic importance that, if realization of the good were burdened by a generally-applicable law, citizens’ support for the regime might be undermined. Many “intimate associations” would likely fall into this category, insofar as citizens value their family and close personal relationships highly and would be unlikely to endorse a regime that burdened those relationships. Beyond this suggestion, I won’t attempt to identify the full range of associations that qualify for exemption consideration under this rubric.

It is important to keep in mind the following: the stability approach does not imply that associations for which exemptions are not warranted are worthy of \textit{no} protection whatsoever, or that all their internal workings ought to be subjected to legal regulation. I would expect that a sensible legal order would refrain from interfering with every kind of association – there is no

apparent public reason to regulate running groups and book clubs. But, in view of associations’
potential destabilizing role, stability does justify reasonable regulation of associations that enter
into the public sphere. Some of these, such as religious associations and some few non-religious
associations, will warrant some special deference. But, in general, legal constraints should be
borne uniformly by individuals, corporate persons, and associations alike. Much more could be
said on this topic, but for now, we will leave it here.

*Noxious Religious Views*

Any discussion of religious toleration must confront the basic problem of noxious
religious beliefs. Some religious persons affirm views, rooted in their religious commitments that
are racist, misogynist, homophobic, or otherwise offensive or antisocial. For our purposes, we
can distinguish two kinds of noxious beliefs. The first are seditious or insurgent beliefs, those
that directly challenge the authority and/or legitimacy of the political order, and perhaps even
call for resistance to it. That these are problematic from the vantage of stability is obvious. The
second category includes offensive beliefs, those that denigrate or demean other citizens.¹²⁸
These beliefs may or may not directly challenge the political order, but undermine conditions of
respect among citizens.

“Unreasonable” religious beliefs have posed a challenge to the stability approach since
we initially defended the reasonableness of stability for the right reasons as a normative ideal.¹²⁹
Indeed, as Jeff Spinner-Halev puts it, the problem of religious beliefs that are resistant to norms

¹²⁸ Note that I have defined offensive beliefs in terms of the subjective sentiments they generate in others, rather than
some objective, content-oriented standard. What matters for stability in regard to offensive beliefs is not their
compatibility with this or that liberal value, say, but rather that they do in fact give offense to others. So here my
focus is slightly different from that of Corey Brettschneider, for example, who is concerned with beliefs that run
¹²⁹ See section 2.3.
of tolerance and political cooperation “haunts liberalism today.” 130 The specifics of religious accommodation are where the rubber meets the road, so to speak. What are the limits (if any) of accommodating these kinds of beliefs in a stable political order? Granting exemptions noxious beliefs on the same terms as for benign beliefs may be undermining to stability in two respects: first, by supporting the persistence of such beliefs and their attendant threats to stability; second, by granting some implicit endorsement to those beliefs (if only by failure to sanction them), potentially magnifying the offense inspired in those whom they demean. All the same, nothing in the argument to this point has provided grounds for discriminating among religious beliefs, and in fact the norm of justificatory impartiality weighs specifically against such discrimination. So what is to be done?

First, the stability approach has no choice but to say that seditious beliefs are to be officially disfavored. This runs counter to justificatory impartiality, and it makes the point where the stability approach veers most sharply away from the core of liberal politics. It is reminiscent of Hobbes’s apparent comfort with state control of religion for the sake of stability, and of Locke’s easy dismissal of toleration for Catholics and atheists on grounds that they cannot be trusted to uphold political justice. 131 There is simply no way around the fact that an argument from stability, without some additional principle of fairness or liberty of conscience, must endorse discrimination against religious beliefs of a seditious or insurgent character.

That said, it is not clear exactly what kinds of practical policies this conclusion entails. Specifically, it may not entail material suppression (i.e., negative distributive partiality) toward persons and associations with seditious beliefs. The reason is that seditiousness is not a binary

condition, and all religious beliefs exist somewhere on that continuum. Consider, for instance, theological statements that cast Christians as “resident aliens” in earthly political communities, or readings that take the Hebrew scripture’s proclamation that “The Lord will reign for ever and ever” (Exodus 15:18) to imply the corollary denunciation, “and not Pharaoh.” Even in light of these statements, I assume most readers would not endorse penalizing or suppressing Christian and Jewish beliefs on the basis of these claims. On the other hand, a recent survey showed that a majority of Americans believe Islam to be “at odds with American values and ways of life,” and major presidential candidates have proposed significant infringements on the civil liberties of Muslim citizens. What this suggests is that judgments about which religious beliefs are seditious or overtly threatening to stability will very often turn on the familiarity of the beliefs in question, rather than on any intrinsic characteristic of those beliefs. Even if stability provides reasons to disfavor seditious religious beliefs, identifying those beliefs that are truly seditious – as opposed to unfamiliar or outside the mainstream – is far from straightforward. In other words, trying to suppress seditious beliefs (by means of force or rhetoric) runs the risk of sweeping up non-seditious persons in its net. In many circumstances, it may be preferable to permit full latitude of belief, on the assumption that the dangers of rooting out seditious belief outweigh the potential gains.

Offensive beliefs pose a different challenge: when granted special accommodation, these pose a threat to stability by giving citizens who are the object of their venom the impression that

---

the political order affirms and supports their tormentors or, in some cases, their oppressors. This can give rise to dignitary harms, as we discussed above. But this would mean that exemptions that grant special status to the expression of offensive views can be ruled out on the basis of third party harms, without requiring a special rule for how to treat specifically offensive beliefs.\footnote{This resolution does rely, however, on a certain violation of justificatory impartiality, insofar as it assumes the wrongness of the offensive belief that those third parties deserve to be denigrated or treated poorly. Without discounting this fact, it is not a problem specifically for my argument here, or for the stability approach in general, but for any political system, which must inevitably “take sides” on some matters.} The principles above can be used to adjudicate exemption requests regardless of whether the motivating beliefs are clearly offensive (racist ideology), clearly non-offensive (drug use in religious ceremonies), or somewhere in between (opposition to abortion rights). Whatever the content of the belief in question, what is decisive for the exemption are the criteria delineated above.

The Supreme Court considered a case involving offensive beliefs in the 1983 case of \textit{Bob Jones University v. United States}, in which the Court upheld the IRS decision to revoke the tax-exempt status of Bob Jones University for denying admission to persons in interracial marriages, and for maintaining a prohibition on interracial dating within the university.\footnote{\textit{Bob Jones University v. United States}, 461 U.S. 574 (1983).} The Court ruled that the government’s interest in eradicating racial discrimination in education was sufficient to justify burdening Bob Jones’s racist religious beliefs. It is not clear, though, that the stability approach supports this conclusion. The IRS policy (of refusing tax exemption to racially discriminating entities) clearly serves a compelling interest, though it’s not clear that application of that interest to Bob Jones University is a very important condition for advancing that interest. Bob Jones is relatively small (less than 3,000 students) and, more importantly, had a history of
wearing its conservative, racially outdated ideology proudly.\textsuperscript{137} Students interesting in the school could be expected to know about this history, and applying was in some respects more akin to joining a social club – or even a church – than patronizing a business or other public institution. To the extent that continuing Bob Jones’s tax exemption imposed a burden on any third parties, it was on the few students interested in attending the school who would have been denied admission. This burden is non-zero, to be sure, but most students would likely avoid applying in the first place. The case is not an obvious one, but from the vantage of stability, it seems unnecessary to revoke the tax exemption for Bob Jones University.

This conclusion is relevant to conflicts currently brewing in and concerning colleges that maintain discriminatory policies in relation to LGBT applicants and students. Some schools forbid students from same-sex dating or from advocating the moral acceptability of sex outside of opposite-sex marriage, and in John Roberts’ dissent from the case legalizing same-sex marriage nationwide, the Chief Justice predicted battles over whether universities would be forced to accommodate same-sex couples in married student housing.\textsuperscript{138} I won’t try to resolve these hypothetical cases here, but I will note that stability might actually lean in the opposite direction than in \textit{Bob Jones} – that is, \textit{against} exemptions for schools with anti-gay policies. This is for two reasons. First, the scope of the \textit{Bob Jones} exemption was narrow: almost no other schools maintained policies against interracial dating. On the other hand, more than seventy leaders of religious schools voiced support for legislation prohibiting the government from

\textsuperscript{137} In 2000, Bob Jones III announced that the University had been wrong to not admit black students prior to 1971, and that it was lifting its policy against interracial dating. See the University’s “Statement about Race at BJU,” available at: \url{http://www.bju.edu/about/what-we-believe/race-statement.php}.

revoking a tax exemption over religious opposition to same-sex marriage. Second, opposition to same-sex marriage is the key issue driving conservative religious activists, and an exemption for religious colleges would be a clear contributor to their political agenda. These considerations lean against an exemption for these schools.

One thing the comparison between *Bob Jones* and same-sex policies highlights is the contingency-dependence of the stability approach. It is impossible to determine *a priori* whether some given exemption should be granted, because some aspects of the question depend on the social and political context. Bob Jones University’s racist policy was, even in 1983, a retrograde outlier, and easy for other citizens to dismiss as such. In contrast, opponents of same-sex marriage in 2015 enjoy wide sympathy, and the struggle over gay rights is very much a live one. These differences matter for stability, and influence the determination of how to handle offensive beliefs.

*Conflict and Conciliation*

Stability for the right reasons relies on the harmonization of loyalties: each citizen judging from within the perspective of her own comprehensive doctrine that it belongs to her good to maintain her support for the regime. In chapter 2, I addressed in the abstract a concern that this aspirational harmonization is unrealistic, illiberal, and/or apolitical. Having fleshed out the constructive argument from the end of this Rawlsian conception of stability, it is worth returning to these critiques. The approach to religious exemptions I’ve advanced in this chapter represents, after all, an attempt to deal with specific conflicts between religious and political obligations, and to formulate the parameters within which these can be harmonized to the benefit

---

of stability. My aim has been to formulate workable principles for reconciling these conflicts, and I recognize that this effort may give the appearance of a naïve hope in reasonableness and the desire of people to “just get along.” Specifically, some of the defenses I’ve offered for the principles above may suggest solutions that are too good to be true: everyone wins, nobody loses. Does this not exemplify the pie-in-the-sky vision that I promised to eschew in rebutting realist objections to the ideal of stability for the right reasons?

There is no denying that the fullest instantiation of stability for the right reasons is a political society in which each person fully endorses their regime. And the achievement of this condition could certainly be described as everyone winning. But it would also entail some losses, as a person’s endorsement of a regime does not imply that it serves their interests perfectly. We endorse regimes all the time that do not maximally reward us: I endorse the rules of the card game even when I lose money; I endorse speed limits even when they inconvenience me on a busy day. Even the ideal of stability for the right reasons, then, does not imply a political society free of loss.

Moreover, I have assumed all along that stability is not a binary condition, and that political societies that are not perfectly stable (for the right reasons) can nevertheless aim at that goal, and can enhance their stability by securing the endorsement of their citizens. Exemptions offer a workable mechanism for pursuing stability piecemeal and incompletely. No single exemption can stabilize a political order, but it can enhance that order’s stability. And remember that the stability approach does not replace the balancing effort prescribed by the RFRA framework, but rather supplements it. The adjudication of each exemption request still depends on a judgment weighing the burden on religious exercise against considerations such as the burdens an exemption would impose on third parties, whether the potential exemptee is
attempting to advance a political agenda, and so on. In virtually any case, some considerations will end up on the losing side of the balance. Some party will be forced to bear a burden on their religious exercise, or some other party will be forced to endure a burden imposed by another party’s exemption, and so forth. Exemptions are, I have claimed, a means by which to enhance net stability, but they are not unequivocally stability-supporting, nor do they serve citizens’ interests without also imposing costs. Applying the framework in real life will, therefore, involve contest and debate, and there is simply no guarantee that perfect harmony will result. Such is the case, of course, with most normative ideals. Yet stability is of sufficient importance that it is worth pursuing net gains where possible, even in full recognition of the necessary tradeoffs.
CONCLUSION:
ASSESSING THE STABILITY APPROACH

6.1 Recapitulation

The argument I have attempted to develop over the preceding five chapters has involved a number of turns, various sub-arguments, and at least a few digressions. A summary is, therefore, in order.

In the closing section of the Introduction, I considered the basic idea of impartiality. While the concept itself is morally neutral, a somewhat more substantive conception of impartiality can describe the core of political liberalism. That conception involves two kinds of impartiality: distributive and justificatory. The first stipulates that the political distribution of burdens and benefits must not depend on the distribution of politically irrelevant characteristics among the individuals constituting the public. The second requires that the political principles prescribing a given distribution of burdens and benefits must not depend on the truth of any particular comprehensive doctrine or doctrines. Taken together, and conjoined to a defeasible presumption that any given individual characteristic is not politically relevant, these two principles of impartiality describe a general political outlook that is consistent with the theories of political liberalism advanced by such authors as John Rawls and Charles Larmore.

In the first chapter, I explored a family of theoretical arguments for the principles of distributive and justificatory impartiality. What united these arguments – which I identified in Adam Smith, J.S. Mill, and John Rawls – is their common invocation of equality as the justificatory ground for impartiality. For each, liberal impartiality is a requirement of guaranteeing conditions of equality among citizens. I then explored how the egalitarian approach to impartiality has been applied to the question of religious exemptions. As it happens, authors
have invoked equality as a consideration both against and in favor of exemptions. Will Kymlicka and Bhikhu Parekh argued for exemptions on egalitarian grounds – the former holding that exemptions equalize the burdens imposed on diverse citizens to maintain membership in the cultural group of their birth; the latter arguing that exemptions are required to secure substantive equality of opportunity. Brian Barry, in contrast, argues against exemptions on grounds that they create inequalities of rights and, implicitly, of citizenship status. What this suggests is that a commitment to equality cannot ultimately resolve the question of exemptions. Exemptions serve inequality in some respects while creating inequalities in others. Some alternative approach to exemptions is needed.

In the second chapter, I defined and defended the conception of stability that would serve as the guiding normative end for the constructive argument that constitutes the core of the project. To set up a contrast with the conception I would ultimately defend, I considered first the idea of ‘modus vivendi’ stability, which I associated with some “realist” critiques of Rawlsian liberalism. I rejected modus vivendi stability, however, on grounds that it specifies an insufficiently robust form of stability; it is persistently vulnerable to the free-rider and mutual assurance problems. In its place, I affirmed a reconstructed version of Rawls’s notion of “stability for the right reasons” (SFRR), which rests on citizens’ reflective endorsement of their regime (and the subsequent likelihood that they will act to support it). I then defended this normative ideal against additional critiques from the realist vantage, including that it is naïve to the challenges of pluralism, or that it presents a basically anti-political vision. I also offered some comments situating SFRR in relation to concepts of feasibility and accessibility and to the ideal/non-ideal theory distinction.
In the third chapter, I advanced a constructive argument defending the general principles of distributive and justificatory impartiality on the basis of the end of SFRR. I relied heavily on insights drawn from Hobbes, specifically observations regarding the various ways in which disagreement and pluralism are prone to generate conflict among citizens. Central to this account were Hobbes’s claims regarding persons’ sense of glory and the potential for disagreement and, more to the point, partiality to give offense. We then explored multiple instances of impartiality in Hobbes’s political thought, including his commitment to the rule of law, to equality, to (qualified) religious toleration, and to some norms of public reason. I argued that these amounted to a conception of sovereign authority that is bounded by the norms of distributive and justificatory impartiality. I then developed arguments along Hobbesian lines defending these two kinds of impartiality impartiality as necessary conditions for stability.

In the fourth chapter, I identified a problem facing impartialist political regimes: namely, the challenge of motivating reliably effective allegiance to abstract, impersonal principles. People are not ordinarily compelled strongly by abstract principles, regardless of their rational justifiability. Even if impartial political arrangements can win citizens’ reflective endorsement, this endorsement risks being ineffective without some additional source of motivation. In reckoning with this problem, we considered three potential solutions. The first was Martha Nussbaum’s account of political emotions, by which the political order would cultivate emotions conducive to support of the regime. The second included Rousseau’s appeals to ethno-national solidarity and civil religion. I identified problems with each account, but the basic weakness shared by both is the inherent difficulty in striking a balance between appeals to abstract ideals (which tend to generate only “watery emotion”) and appeals to particularistic creeds and identities that risk becoming exclusive, homogeneous, and incompatible with pluralism. In place
of these options, I defended what I called a decentralized approach that drew on concepts from John Locke and John Rawls. In the former’s *Letter Concerning Toleration* and the latter’s idea of the overlapping consensus, allegiance to impartialist norms is secured by appeal to citizens’ respective (and diverse) comprehensive doctrines. I argued that this approach is both reasonable and more likely to be effective than either Nussbaum’s or Rousseau’s.

The fifth chapter returned to the problem of religious exemptions. On the basis of the account of liberal impartiality from the end of stability, I formulated a framework for treating religious exemptions that is conducive to stability for the right reasons. In general, stability requires a norm of impartiality, establishing a presumption against differential treatment on the basis of religious belief. However, I argued, there are two ways in which stability may actually support exemptions for religious practice. The first is in order to avoid subjecting citizens to burdens that they cannot regard as justified. When laws and institutions infringe upon citizens’ religious practices without a good reason, those individuals feel disadvantaged. Selective exemptions can ease this subjective sense of disadvantage and shore up those citizens’ support for the regime. Second, insofar as citizens’ diverse beliefs and commitments can serve as resources to overcome the problem of ineffective endorsement, exemptions targeted at supporting the flourishing of diverse moral communities can indirectly serve stability. We compared this approach with the reigning legal doctrine governing exemptions in U.S. law, and considered its implications for a number of cases, including the recent *Hobby Lobby* case. I concluded by addressing some potential objections to the approach and giving some indications of what it might recommend in future cases.
6.2 Some Elements of Significance

In the Introduction, I suggested that the merits of the stability approach, whatever they might be, would be made known by its fruits. That is, I did not attempt to defend the stability approach in some a priori manner, but rather hoped that by pursuing the line of argument, its strengths (and weaknesses, of course) would become apparent. We are in position now to assess some of those strengths. We will consider the weaknesses momentarily.

The first, and most straightforward, accomplishment of the project is that it provides a coherent normative framework for addressing the challenge of religious exemptions. That is, the stability approach offers a cohesive set of considerations that can be employed in determining whether, and when, an individual’s religious belief or religious identity warrants a dispensation from some legal obligation. We tried to work out some of the applications of this framework to actual cases in Chapter 5, and saw that the stability approach can ground judgments regarding the desirability of exemptions in particular circumstances. This is not to say that it is perfectly determinate: stability is, in some respects, ambivalent on the question of exemptions. In this respect, the stability approach and the egalitarian approach are similarly incapable of providing unequivocal and conclusive conclusions in each and every case. But even while the application of the stability approach to concrete cases is complex, it is not impossible. Rather, identifying the stability-supporting conclusion is a matter of balancing considerations against one another on terms that are conceptually commensurable.\(^1\) Moreover, the value of the stability approach does not depend on it being the only – or even the best – possible such approach. Instead, it offers meaningful normative guidance on questions where other existing normative frameworks fall

---

\(^1\) A more significant challenge lies in prediction: assessing what the consequences to stability, as defined by the parameters identified in chapter 5, are likely to be. The stability approach, as a consequentialist framework, will be always troubled by the fact of uncertainty and the difficulty of forecasting. But this is not unique to the stability approach, nor is it entirely damning.
short of decisive resolution. This guidance can serve to both inform and critique judicial, executive, and legislative policymaking on religious exemptions.

The normative framework supplied by the stability approach has the additional merit of being fully freestanding of substantive moral commitments. Whereas the equality approach depends upon the judgment of equality (however conceived) as a moral good worth pursuing, the stability approach does not rest on any substantive moral premises. Instead, it assumes the practical good of political stability, as compared to instability. In the era of post-Rawlsian (perhaps even post-Nietzschean) liberalism, much has been made of the purported problem of justifying liberalism on terms that are themselves neutral. I do not intend here to take a stance on whether political theory must, or must strive to be, freestanding “all the way down.” Rather, I wish only to appeal to those who do endorse this view, or some version of it, and draw attention to the fact that the stability approach offers promise in this regard.

I have explicitly framed the normative framework as an extension of the Hobbesian project, which sought to derive political principles from the practical good of order. Just as Hobbes’s argument, as I can best grasp it, remains a conditional one – one must endorse an unlimited sovereign only if one desires the safety that comes with political order – so too the argument I have advanced here. I am confident, however, that the foundational assumption of stability’s desirability is widely shared. All of this said, it is worth noting the possibility that the argument depends crucially on some moral assumptions which I have not acknowledged. I have tried scrupulously to ferret out any such “inputs,” and I leave it to some other critic to identify my shortcomings in this regard.

---

2 Stability may also be morally preferable to instability, but my argument does not depend on this being the case. I did address briefly, in chapter 2, the view that instability is morally superior to stability insofar as such conditions are generative of human creativity, strength, and excellence. My position is that the practical disadvantages to instability are sufficient to outweigh whatever moral good it may produce.
A third credit to the argument is that it elaborates and extends existing assessments of the liabilities and possibilities of pluralism. In chapter 3, we followed Hobbes in recognizing the “perils” of pluralism, including the conflict-generating potential of disagreement and group associations. The constructive argument advanced there was explicitly aimed at mitigating those liabilities by means of smart political principles. In chapter 4, we considered another face of pluralism, showing that a rich diversity of moral communities can be indirectly stability-supporting. Pluralism performs this function by providing each citizen with some deep commitments that can motivate their adherence to political norms of impartiality. To be sure, this fact coexists with the facts of pluralism’s potential to undermine stability, so the observation should be not be interpreted as unduly sanguine about the effects of pluralism. The observations in chapters 3 and 4, taken in conjunction, add up to a nuanced theoretical estimation of how pluralism can both challenge and support political stability. While this ambivalent stance complicates the application of the stability approach to particular cases, it honestly accounts for the mixed bag that pluralism contributes to liberal political life. With this argument, I have attempted to both bridge and contribute to two bodies of literature: those that plumb Hobbes for insights into the challenges of pluralism, on the one hand, and those that look to civil society for resources to support and enrich liberal political life, on the other. Other authors, such as Jacob Levy in his recent *Rationalism, Pluralism, and Freedom*, have accounted for historical ambivalence toward pluralism in the liberal tradition. My contribution is to draw out this complexity in normative terms.

A fourth accomplishment of this project is narrower in scope, yet significant in scholarly terms all the same. In advancing a constructive argument from the foundational end of stability, I have attempted to indirectly contribute to scholarship on Rawls that has emphasized the role of

stability in his theory of justice. In the second chapter, in particular, I drew substantially on Paul Weithman’s 2010 book Why Political Liberalism?, a major work of interpretation that persuasively argues for a reading of Rawls’s “political turn” in which stability figures centrally. It was because Rawls doubted the tenability of the argument for the stability of justice as fairness in A Theory of Justice that he recast parts of the argument in Political Liberalism. My project has not been primarily exegetical, nor primarily focused on Rawls, but I have drawn inspiration from Weithman’s work and offer my argument in a spirit of conversation with his interpretation of Rawls. Insofar as Weithman has suggested that the priority of stability is responsible for certain aspects of Rawls’s mature work (such as the overlapping consensus and the idea of public reason), one is tempted to wonder what other elements of a liberal political theory might be traceable in some way to concern for stability. While I do not mean to impute anything like the stability approach to Rawls himself, I have tried to show that some basic elements of a political liberalism can be derived from the end of stability alone, without invoking additional goods such as equality or freedom. In the third chapter, I wondered whether there might be something Rawlsian about Hobbes’s conception of stability; one way to understand the argument of the dissertation is that I’ve argued that major components of Rawls’s mature views can be understood and defended in Hobbesian terms by taking (a certain conception of) the end of stability as normatively prior and deriving principles of political liberalism from it.

More generally, the project points toward an interpretive reimagining of the liberal tradition as a whole. Just as the argument suggests taking the normative end of stability more seriously as a component of Rawls’s theory of justice, it also suggests taking stability more seriously as a driving consideration in liberalism more broadly. I am not making what would be an obviously erroneous and overreaching claim: that stability is what liberalism is all about.
Rather, I am proposing that accounts of liberalism attend more thoroughly to the end of stability in both the historical experience of liberalism as well as its philosophical defenders. Not only Hobbes but also Locke inaugurated the liberal tradition with contract doctrines aimed at resolving the consequences of disorder implicit in the state of nature. While most accounts of liberalism (including Locke’s) have been motivated by moral priorities as well (such as rights protection, equality, fairness, etc.), the end of stability has often lurked nearby, if often just out of sight. There is not space to pursue this suggestion at length here, and in any case my project is not principally a historical one. But insofar as my argument has demonstrated the normative potency of (a certain conception of) stability, it may be fruitful to attend more closely to the normative role stability has played in the history of liberalism. Prospectively, the argument offers a first step in a more comprehensive reformulation of liberalism on the basis of stability. At the very least, it suggests the possibility that an effort in this direction may be worth pursuing.

6.3 Caveats and Limitations

Having noted these several points to the credit of the argument I have developed here (some of which may include a degree of overstatement), it is only appropriate to note some of the argument’s limitations as well. The points that follow, it should go without saying, stand alongside those flaws and oversights of which I myself am not aware. Even from the decidedly partial authorial vantage, though, it is possible to note a few of the project’s limitations.

First among these is the simple fact that the normative framework for dealing with exemptions is so complex – by which I mean that it involves multiple, irreducible elements – as to preclude neat and definitive resolution of most cases. As the examples in chapter 5 indicated, tracing an easy route from the end of stability to the proper adjudication of some particular
exemption request is usually impossible. The reason is that stability is profoundly ambivalent on the accommodation of difference, with considerations leaning in both directions. In any particular case, one is left to balance these competing considerations against one another in hopes that the tipping of that balance will be possible to discern. For many potential exemptions, however, there will be no decisive consideration either for or against, and the ultimate decision will be a matter of judgment. This does not preclude the stability approach from being used profitably by persons in position of authority to grant or deny exemption requests, but it falls short of supplying incontrovertible answers in each and every case.

Second, though I have advanced this project under the heading of liberal impartiality, the constructive account does not imply any guarantee that political arrangements derived from the end of stability will be perfectly liberal. We have noted this fact in a couple of places already, but it bears repeating, insofar as many readers will reasonably wish to dismiss any political theory that fails to provide decisive grounding for core liberal commitments. The most that the stability approach can guarantee is that the basic structure of a stable political regime will be capable of winning the reflective endorsement of the citizens living under it. This means that profoundly discriminatory, abusive, or arbitrary regimes will be precluded, and the norms of distributive and justificatory impartiality are likely to ground specific political principles that accord with liberal sensibilities. However, there is no way to be certain, a priori, that the principles a given public can accept will not contain illiberal elements. For example, in a society with strong norms of piety, sensitivity, and discretion, censorship or anti-blasphemy laws might be stable. Or, the economic arrangements sanctioned by distributive impartiality might range from robustly free-market capitalism to (limited) central planning in the mode of strong democratic socialism. While I have argued that the stability approach grounds a political paradigm that is broadly
consistent with the norms of political liberalism, it does not guarantee fully liberal laws and institutions. This may strike committed liberals as a serious shortcoming.

Third, while I have made a point of emphasizing that I do not intend the end of stability to necessarily displace other substantive political ends (including moral ones), this leaves open the possibility of conflict between stability and these other ends. Recall that Rawls’s work sought political principles that satisfied two ends: justice and stability. He wanted a political framework that could be “stably just.” In this project I have attempted to isolate the normative end of stability, but I have not insisted that other normative ends are to be rejected or dismissed. Indeed, I have just affirmed that a reliably liberal regime cannot be guaranteed by stability alone – some additional end may be required. These other ends (e.g., fairness, justice, equality, good-promotion, etc.) may be to some degree compatible with stability, but they may also conflict. At the very least, showing that some set of political principles can satisfy multiple ends requires a degree of effort that is not required to show that those principles satisfy one end, such as stability. One goal of this project might have been a comprehensive political theory that situates stability as either dominant over or compatible alongside other normative ends. This particular goal remains unmet.

Fourth, the account I have advanced relies on a particular moral and social psychology – one inspired substantially by Hobbes. The details make up the bulk of chapter 3. In short, we accept two premises about human motivation: self-interest and the desire for honor. The first, for our purposes, amounts to a claim that in order for a person to support a regime, she must believe that the regime serves her interests. The second asserts that people take offense at perceived slights to their status (what Hobbes calls ‘glory’), and that they are less likely to support a regime that offends their sense of honor. A third premise, which Hobbes would also affirm, is that

---

religious beliefs (and potentially some other kinds of beliefs or commitments) are capable of motivating persons with sufficient strength as to override other sources of motivation. I have proceeded on the assumption that these premises are minimal, and have tried to offer the most plausible presentation of them that I can. But it remains that if one denies these premises – or if one thinks that there are additional motivational factors I have not accounted for – then one may be unpersuaded by the remainder of the argument.

Fifth, the conception of stability that I have employed throughout the project – stability “for the right reasons” (SFRR) – is highly particular, and may be of limited use in actual contexts of instability, disorder, and conflict. In the second chapter, I addressed the objection that SFRR describes a kind of stability that is so unlikely to be realized that to adopt it as a normative ideal reflects a sort of naivete. I defended the proposition that while SFRR sets an ambitious standard for stability, it is a thoroughly reasonable one, with no reason in theory or practice that it could be not be attained. This is not the same, however, as suggesting that SFRR is a reasonable or realistic or proper goal for any and all actual political contexts. Such a proposition cannot be defended. In zones of outright military conflict, for instance, the appropriate end ought to be mitigation of violence and the establishment of basic order. One thinks, for instance, of some regions in the Middle East that are currently objects of contest between state governments and non-state actors such as ISIS, or of countries like the Democratic Republic of the Congo, which has suffered under civil war for nearly two decades. In such contexts, the immediate aim must be establishment of the peace, even if by means other than such would be consistent with SFRR. The ideal of SFRR is most practically useful in contexts where there already exists some kind of a modus vivendi. This project, by focusing on questions concerning religious accommodation in
the U.S., has stayed safely within the bounds of such a context. Determining the precise location of those bounds is a separate task entirely.

There are, undoubtedly, many other limitations, weaknesses, or outright failures of the project, in addition to the ones I have identified here. In highlighting the five items above, I do not mean to give the impression of hubristically believing that I have addressed all potential objections. Rather, I hope only to note some of the limitations of the project, if for no other reason than to be clear about what I am and am not claiming can be established on the basis of stability.

6.4 Future Trajectories

This project has aimed specifically at addressing the issue of religious exemptions, but it has done so under the heading of broader notions such as stability, pluralism, and impartiality. The argument from stability is not limited in application to matters of religious accommodation. Looking ahead, there are a number of ways in which the argument might be extended to address other outstanding questions in liberal theory. Here I want to identify just three such questions, and to offer some preliminary commentary on how the stability approach might offer novel insights into these additional areas of inquiry.

Political Education

Pluralist societies are inherently vulnerable to fracture and disunity. Citizens who affirm diverse, often incompatible beliefs may locate their primary identities in relation to those beliefs, rather than in relation to their status as co-nationals or members of a common political society. This presents several challenges. First, on some views, social cohesion and political unity may be
valuable in and of themselves. Second, disunity troubles the promotion of certain values as elements of a shared identity or ethos. Third, disunity, insofar as it diminishes citizens’ identification with and likelihood of supporting their political society, is potentially detrimental to stability. One of the key mechanisms that a political order has at its disposal to promote social unity is the public education system. This is the principal means by which politics can intervene in human development and contribute to the formation of burgeoning citizens. A system of education can facilitate assimilation, transmit norms and values, foster shared identities, and encourage cohesion.

Pursuit of these goals, however, is bound to run up against issues concerning the proper degree of deference to pluralism. Shared identity is a good, but, as I’ve argued in the preceding chapters, it is both an impractical and an illiberal goal to entirely suppress meaningful differences in the aims and commitments that citizens adopt. Some middle course – between social disintegration and imposed uniformity – must be found for the case of education. Given this dynamic, it should be no surprise that a significant portion of American jurisprudence concerning religion and religious exercise stems from issues arising in schools. To name just a few: the permissibility of prayer (voluntary or compulsory) in public schools; the permissibility of religious instruction in public schools; the permissibility of providing state funds for private religious schools; whether families may opt out of compulsory education for religious reasons; and so on.⁵ The relevant issues, though, are not limited only to religion. The same issues could be raised in relation to other kinds of deep commitments (though they would not all raise the same constitutional questions as religion). As Martha Nussbaum has explored in sections of her book *Political Emotions*, the basic question is how a society may, by means of the education system, cultivate both enthusiasm for a regime and allegiance to it while 1) respecting individual

autonomy and maintaining at least some degree of neutrality among citizens’ deep commitments, and 2) ensuring that allegiance to the political order does not become entirely uncritical, and consequently pathological.⁶

The stability approach is well-equipped to make a contribution to these fields of discourse. The questions implicated in the context of education are highly analogous to – and in some cases identical with – the questions implicated in the area of religious accommodation. Most basically, stability is served when citizens grow up to be supportive of their regime. The stability approach is positioned to formulate normative parameters for political education as it did for religious exemptions. Some additional work is required, though, to determine at least the outlines of the content of political education. In the religious exercise context, we did not need to deal with the content of political principles or laws – the focus was on the content of citizens’ beliefs. In education, however, the content of political education must be determined, and the stability approach will need to contribute to this determination.

Public Reason

Since the initial articulations of political liberalism by Rawls and Larmore in the 1980s and early 1990s, the ideas of public justification and public reason have come to enjoy a place of prominence in contemporary liberal theorizing. “Public reason liberalism” has been described as “perhaps the dominant brand of liberal political theory today.”⁷ Debates concerning public reason have focused on both the justification(s) for the rules of public reason and the content of those rules. In particular, a great deal of ink has been spilled over the admissibility of claims

---

based on religious beliefs in fora of public reason. Since Rawls’s call for “freestanding” political proposals (later modified by the “proviso” permitting limited appeals to religious beliefs), some authors, such as Robert Audi and Steven Macedo, have affirmed the exclusion of religion from public reason, while others, such as Nicholas Wolterstorff, Christopher Eberle, and Kevin Vallier, have argued for the inclusion of religious beliefs. The debate remains a live one, though recent energy appears to have tilted the balance, at least for now, in favor of the inclusive view of public reason.

Our investigation of stability in this project has approached the public reason debate in at least two instances. First, in chapter 2, I defended an interpretation of stability for the right reasons that relied on a “convergence” of citizens’ reasons for supporting their regime. This term – which I counterposed to a “consensus” view – is drawn from the public reason debate: the inclusion of religion in public reason can be defended as supporting a convergence of justificatory reasons. Second, in chapter 3, I defended the principle of Impartiality of Justification, which implies that political arrangements may not rely for their justification on any particular comprehensive doctrine (including religious ones). This sounds suspiciously like an endorsement of the exclusive view of public reason. That the stability approach has engaged with the categories of public reason both indicates that it may be able to fruitfully intervene in the public reason debate and suggests that such an intervention may be a necessary step in resolving an apparent inconsistency in the argument thus far.

Carrying out such an intervention will require some further refinement of the norm of Impartial Justification, as well as further discussion of the specific characteristics of justifying reasons that qualify or disqualify them from inclusion in public justification. But we can offer at least one speculative comment at this point. Note a distinction between the defense of
convergence in chapter 2 and the norm of Impartial Justification. Chapter 2 dealt with the conditions under which a political regime can be stable, and convergence was invoked in support of my claim that citizens do not need to agree on the reasons for a regime’s justification. Rather, it can be stable if they each recognize some reason, even if non-shared, to endorse the regime. The norm of Impartial Justification, in contrast, deals with the motivational consequences of a citizens’ recognizing that a political arrangement assumes the truth of some premise they deny. Thus, convergence of justifications is sufficient for stability when considered from the vantage of each individual citizen in relation to their regime, but it may not be sufficient in the social exchange of reasons in public deliberation. Defending this claim in more detail requires much more space than is available here, but the basic suggestion is that the stability approach may contribute to public reason debates by distinguishing the fact of justification from the project or performance of justification.  

Defending Liberal Impartiality

In the Introduction, in the process of specifying a conceptually parsimonious account of impartiality, we took care to distinguish the norm of impartiality itself from any purported moral value inhering in impartiality. Impartiality is itself morally neutral, and affirming this means avoiding the temptation to equate impartiality with morality. This temptation is not universally felt, however. Certain ethicists and political philosophers have argued the opposite: that impartiality is itself the problem at the root of many socio-political injustices. For example, Iris

---

8 There is another salient distinction, which is that IJ (as I have formulated it) does not insist that political arrangements be based on shared reasons (as some defenders of exclusive public reason, such as Macedo, would hold), but rather that they be justified by reasons that are not dependent on any particular comprehensive doctrine. This would suggest that a reason need not be fully shared by the entire political community so long as it does not depend for its truth on any one comprehensive doctrine. Thinking through these norms of impartiality in the context of public reason may shed light on the tenability of this formulation.
Marion Young offers a compelling attack on the norm of impartiality in her 1990 book *Justice and the Politics of Difference*. Young critiques impartiality as reflecting an unrealistic aspiration to a universalist perspective in reason, and as misguided valuing the general as more morally important than the particular. The tragic consequence of the aspiration to impartiality in politics, on Young’s diagnosis, is that the law is rendered blind to important differences between people, and thus while formal equality may be achieved, substantive equality lacks. This objection to impartiality is echoed in a number of feminist critiques of liberalism. From a different direction, Bernard Williams has critiqued impartiality in moral theory, famously suggesting that in many cases of moral reasoning, stopping to consider the perspective of impartiality may constitute “one thought too many.” Similarly, Charles Taylor has pushed against the notion of a rational perspective that is divorced from a context in which it might be situated, and which is disconnected from the “constitutive goods” that give shape to human lives and human reason.

Or, some theorists such as Bonnie Honig and William Connelly who embrace an “agonistic” conception of politics have suggested that liberal impartiality is inhospitable to the flourishing of genuine pluralism. Whereas I have tried to show that the stability approach grounds norms of impartiality that are generally compatible with and supportive of the aims of justice, these critical approaches allege that the goal of justice is not well-served by impartiality.

---

I have defended a certain kind of impartiality, but I have not suggested that political arrangements must be impartial in an unqualified way, though. Rather, stability can ground sensitivity to difference in some cases. Nevertheless, the stability approach sets impartiality (both distributive and justificatory) as the presumptive norm for liberal politics. The stability approach may therefore be both vulnerable to the aforementioned critiques of impartiality and also equipped to address those critiques in ways not possible on the terms of egalitarian accounts of impartiality. At this point, I am not prepared to offer a full analysis of and response to these critiques, but can offer a few speculative comments. First, it is my suspicion that at least some opponents of impartiality impute to political conceptions of impartiality requirements that individuals strive for impartiality in their private reasoning and judgments. This, as I’ve stressed throughout the project, is not required by the stability approach. Second, the stability approach may be uniquely capable of resisting, defusing, or sidestepping the anti-impartiality critiques. For one thing, the stability approach grounds deviations from impartiality in cases where impartiality imposes undue burdens on individuals or groups. For another, the stability approach will shed helpful light on the stability-undermining effects of too much sensitivity to difference.

I offer each of these possible areas of future development only provisionally; my attempts to outline both the important questions and my prospective answers require much more thought before I am prepared to stand behind them. However, they are suggestive of the possibility that the stability approach might have application well beyond the context in which we have developed it here.
6.5 Religion and Liberal Politics

In a couple of places throughout this project I have asserted that, though the specific matter at issue is the question of religious exemptions, religion is not, *per se*, the object of my interest. Religion is, I’ve held, a paradigmatic example of a comprehensive doctrine, and religious disagreements are paradigmatic examples of reasonable disagreements among citizens. Religion stands in nicely for the broader set of moral, epistemic, and practical commitments that populate the psyches of a citizenry. Focusing on religion is, I’ve suggested, a move for the sake of narrowing down the subject matter and simplifying the argument.

This is only partially true. Religion does exemplify the dynamics of pluralism that are salient for the arguments we’ve considered above, and zeroing in on religion does help to distill those dynamics so as to be more conceptually manageable. But religion also holds particular interest for me personally, and, more to the point, I would contend that there are good reasons to lend our attention to religion itself, not merely as a species of comprehensive doctrines or pluralism more generally.

Religion occupies an uncertain place in liberal politics, and always has. Liberals tend to divide over their estimations of religion, its merits or value, its social effects, and its proper place in liberal political society. This includes empirical questions: after centuries in which the “secularization thesis” predominated among social scientists, recent authors have called into question the assumption that religion would wither away in modernity.14 With apologies to Charles Taylor, it seems ever more clear that we are living in a *post*-secular age.15 The fact seems to be that people are inalienably drawn to religious beliefs – in myriad forms. For good or bad, the religious impulse is a deeply rooted feature of the human experience, both individual and

---

social. At the same time that the empirical thesis of secularization is waning in influence, normative theorists are pushing back against the marginalization or outright exclusion of religion in liberal politics. Kevin Vallier has distilled these objections to two basic arguments: 1) that the relegation of religion to the private sphere forces religious citizens to sacrifice personal integrity by leaving their religious convictions “at the door” of public life; and 2) that the marginalization of religion unfairly disadvantages religious citizens, where non-religious citizens are not required to sacrifice or obscure their deepest convictions.\(^\text{16}\) As mentioned in relation to the public reason debate above, the notion that religion must be sidelined in liberal politics has fallen under significant attack.

It is not the case, however, that liberals have come to embrace religion without exception, or in an unqualified way. The work of so-called “new atheists,” such as Sam Harris, Richard Dawkins, and Christopher Hitchens, represents a popular wave of disdain for religion on both epistemic and moral grounds.\(^\text{17}\) Not only are many of the commitments often associated with religion (theism, revelation, life after death) epistemically unjustified, these authors argue, but they are also responsible for all manner of objectionable behavior, including the Crusades, suicide bombings, and homophobia. These authors are unabashedly bombastic, and some of their arguments lack sophistication, but many of their basic claims are shared by more serious scholars. Brian Leiter’s recent *Why Tolerate Religion?*, for instance, relies on the claim that religious belief involves “culpable failures of epistemic warrant.”\(^\text{18}\) And even those who


champion religion recognize its potential for ugliness. Additionally, some liberals worry that religious belief, and/or membership in religious communities, is incompatible with a fully developed sense of autonomy.

The uncertain status of religion in liberalism extends to the realm of actual politics. At present in the United States, loud voices raise alarms decrying a supposed “war on religion,” or announcing that religious liberty is “under attack.” At the same time, others warn of a slow creep toward theocracy. Of the two major political parties in the U.S., one loudly affirms (Christian) religious belief and relies deeply on the votes of religiously-motivated citizens, while the other curries more favor with the less-religions and has struggled in recent years over its relationship to religious voters. The 2016 presidential campaign has seen a robust (if hyperbolic) debate among Republicans over the compatibility of a particular religious tradition, Islam, with public safety and liberal-democratic values. Religion is both central to American collective self-identity and, at least in some corners, deeply suspect.

A posture of ambivalence toward religion in public life is entirely justified, given the mixed contributions that religion has made to world history. On one hand, religion has motivated much charity, grounded much solidarity, and provided much solace. At the same time, one need not be a full-blown Dawkins or Harris to acknowledge some horrific evils perpetrated under the

guise of religion. The annals of religious action include both Mother Theresa and Osama bin Laden, Mohandas Gandhi and David Duke. Religion births both heroism and villainy, compassion and antipathy, works of love and acts of war. There simply is no way around what Scott Appleby calls “the ambivalence of the sacred.” Moreover, it is impossible to make some definitive reckoning to determine whether religion has been a net positive or a net negative in human history. The causal links are too attenuated and/or complex, and the consequences are of incommensurable kinds.

We are left to wonder whether there is some way to harness the good of religion while resisting, suppressing, or at least containing the bad. One way to do this is from within a particular religious tradition, by means of efforts in theology, political theology, theological ethics, ecclesiology, missiology, etc. These are confessional projects that determine the shape that a religion’s public presence will take. Ordinarily, persons internal to that religion – i.e., members of the faith – will be the ones engaging in battle over diverse interpretations or orientations. Relatedly, there is a temptation for persons external to a tradition to intervene in these areas, tinkering with the internal beliefs and dynamics of religious traditions in order to ensure that they are suitable for liberal-democratic contexts. I include as part of this phenomenon government efforts to identify and promote “moderate” forms of Islam. The arguments above in favor of Justificatory Impartiality would weigh against attempts to “take sides” on theological or other substantively religious matters. In the American context, the Establishment Clause also provides a strong constitutional norm to the same effect. What’s more, outside determination of questions internal to a religion tends to result in an ugly hypocrisy, whereby unfamiliar faiths are subjected to a public scrutiny that mainstream faiths are not, even though each contains the

25 Angel Rabasa et al., *Building Moderate Muslim Networks* (Santa Monica, CA: Rand Corp., 2007).
potential for both virtue and havoc. (See public insistence that Muslims denounce acts of terrorism committed in the name of Islam, while no such demands are made of Christians.) For these reasons, theological questions are best left to the theologians.

What, then, is left to do for those of us who are concerned not with this or that religion, but with religion itself? Given that religion will take all kinds of different forms, what can the political order do to promote and reap the benefits of religion while minimizing/mitigating the dangers it presents? One possibility, as I’ve noted in a few places, is to suppress religion by means of force. This approach, of course, is not acceptably liberal, and I doubt that it is practically feasible. The stability approach that I have defended is an attempt at a properly political (as opposed to theological, etc.) approach to religion as an ambivalent social force. I have aimed to treat religion as a category, while also remaining sensitive to religious diversity. Taking this approach has affirmed that religion – both in its distinctiveness and as a species of pluralism – represents both an asset and a liability to liberal society. The normative approach I’ve recommended affords substantial latitude to religious beliefs, communities, and practices to flourish. I have not recommended political intervention into religious doctrine or practice, and I have recommended against most attempts to discriminate among religions on the basis of the content of their beliefs. There are a number of good reasons for liberals to affirm this as the best approach, but I hope that I’ve shown that stability provides distinctive reasons of its own.

One inevitable consequence of this approach is that some religions that are permitted to flourish will produce net negative consequences for the political society. If the law is not empowered to suppress “bad” religion and promote “good” (however those descriptors are defined), it cannot guarantee that some religions will do more harm than good. The stability approach accepts this fact.
We can invoke a religious image to characterize this stance. The Gospel of Matthew relates Jesus’s parable of fieldworkers who asked their employer whether they should root out weeds growing amid the sown wheat. The head of the household says no: by pulling out the weeds, they would also uproot the good wheat surrounding them. Instead, he directs them, “Let them grow together until the harvest.” The argument from stability I’ve advanced here ultimately recommends that liberal politics permit diverse forms of religion to “grow together.” This is not to take an unreasonably sanguine view of religion’s perils, but rather to note that to suppress or constrain it would mean precluding the good it may provide. My optimism on this count reflects some assumptions that haven’t been explored here, including a degree of confidence in the ability of democratic deliberation to marginalize truly nasty beliefs, as well as my personal conviction that a diverse landscape of religious beliefs and practices enriches the existences of individuals, collectives, and society as a whole beyond what can be tabulated in material terms.

This is an optimism, to be sure, but I hope that the argument presented here provides some support for maintaining it. Liberal politics need not choose between difference and cohesion, religious conviction and social order, pluralism and stability. On the contrary, while the dynamics are by no means simplistic or free of risk, a society rich in diversity can indeed be just, good, and stable.

---

BIBLIOGRAPHY


Sleat, Matt. “Realism, Liberalism and Non-Ideal Theory Or, Are There Two Ways to Do Realistic Political Theory?” *Political Studies*, 2014.


Court Cases:

Fraternal Order of Police Newark Lodge No. 12 v. City of Newark, 170 F.3d 359 (3d Cir. 1999).
Reynolds v. United States, 98 U.S. 145 (1878).
Thomas v. Anchorage Equal Rights Commission, 220 F.3d 1134 (9th Cir. 2000).
United States v. Quaintance, 608 F.3d 717 (10th Cir. 2010).
United States v. Zimmerman, 514 F.3d 851 (9th Cir. 2007).