FAITH IN THE BALANCE? AN EVALUATION OF THE OUTCOMES OF
RELIGIOUS FREE EXERCISE CLAIMS IN THE STATE APPELLATE COURTS,
1997-2011

A Dissertation in
Sociology
by
Robert R. Martin

© 2013 Robert R. Martin

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2013
The dissertation of Robert R. Martin was reviewed and approved* by the following:

Roger K. Finke  
Professor of Sociology and Religious Studies  
Dissertation Adviser  
Chair of Committee

Duane F. Alwin  
McCourtney Professor of Sociology and Demography

John D. McCarthy  
Distinguished Professor of Sociology

Christopher J. Zorn  
Liberal Arts Research Professor

John Iceland  
Professor of Sociology and Demography  
Head of the Department of Sociology and Criminology

*Signatures are on file in the Graduate School.
ABSTRACT


Religious believers across the globe have two sets of obligations: the obligation to follow the dictates of their creed and the obligation to obey the law of the state. Though these obligations often exist in harmony with one another, they come into opposition when conduct mandated under one set of obligations is proscribed under the other. The federal Constitution and all state constitutions protect the right to practice one’s religion freely, but how durable are these protections in the face of laws that prohibit what believers feel they must do because of religion? This dissertation explores the results of the judicial process Americans pursue when their religious and legal duties come into conflict, and does so in the context of the state appellate judicial system. It uses multivariate statistical modeling techniques to answer two related questions: (1) what factors predict whether a free exercise claim will be upheld or rejected in the state courts, and (2) what factors predict the form of legal test on which judges base their decisions. Legal positivist and legal realist frameworks from the sociology of law are employed to address these research questions. Content analysis of published opinions in the state appellate court systems from mid-1997 to 2011 was conducted to generate the dataset employed to test hypotheses related to the two research questions. The dissertation finds that claimants in state appellate free exercise cases are rarely successful; the vast majority of claims are not upheld by the courts. The dissertation also finds that when judges use legal tests that balance the severity of burdens imposed by the government on religious exercise against the government’s interest in doing so, they are more likely to uphold free exercise claims than when a rational-basis test is
used. Unlike previous studies of free exercise disputes in the courts, however, the dissertation
does not report differences in case outcomes based on the religious affiliations of claimants.
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables ........................................................................... vii</td>
</tr>
<tr>
<td>List of Figures ........................................................................... viii</td>
</tr>
<tr>
<td>Acknowledgments ........................................................................ ix</td>
</tr>
<tr>
<td>Chapter 1. INTRODUCTION AND THEORETICAL BACKGROUND .......... 1</td>
</tr>
<tr>
<td>Introduction .............................................................................. 1</td>
</tr>
<tr>
<td>The Courts and the Meaning(s) of Free Exercise Rights ............... 3</td>
</tr>
<tr>
<td>Theoretical Frameworks: Legal Positivism and Legal Realism ........ 6</td>
</tr>
<tr>
<td>The Consequence of Decisions in Free Exercise Cases: Implications from Religious Economies Theory ........................................ 13</td>
</tr>
<tr>
<td>Summary and Organization of Dissertation ................................ 18</td>
</tr>
<tr>
<td>Chapter 2. FREE EXERCISE RIGHTS – HISTORY AND HYPOTHESES .... 22</td>
</tr>
<tr>
<td>The Courts and the Free Exercise of Religion: A Primer ............... 22</td>
</tr>
<tr>
<td>I. Early Free Exercise Jurisprudence: 1813-1963 ......................... 23</td>
</tr>
<tr>
<td>II. The Rise of the Compelling Interest Test: 1963-1990 ............... 25</td>
</tr>
<tr>
<td>III. The Contest Over the Meaning of Religious Free Exercise: 1990-1997 ................................................................. 27</td>
</tr>
<tr>
<td>IV. City of Boerne v. Flores and a Patchwork of Meanings: 1997 to 2013 ........................................................................... 29</td>
</tr>
<tr>
<td>Review of Free Exercise Research ............................................... 32</td>
</tr>
<tr>
<td>Dissertation Hypotheses ............................................................ 39</td>
</tr>
<tr>
<td>I. Positivist Hypotheses .............................................................. 40</td>
</tr>
<tr>
<td>II. Realist Hypotheses ................................................................. 44</td>
</tr>
<tr>
<td>Summary .................................................................................... 47</td>
</tr>
<tr>
<td>Chapter 3. DATA AND METHODS ................................................ 49</td>
</tr>
<tr>
<td>Content Analysis of Court Decisions .......................................... 49</td>
</tr>
<tr>
<td>I. Using LoislawConnect to Locate Decisions ................................ 50</td>
</tr>
<tr>
<td>II. Case Selection ....................................................................... 52</td>
</tr>
<tr>
<td>III. Measurement Reliability ....................................................... 55</td>
</tr>
<tr>
<td>Measures .................................................................................. 58</td>
</tr>
<tr>
<td>I. Dependent Variables .............................................................. 58</td>
</tr>
<tr>
<td>II. Independent and Control Variables ....................................... 62</td>
</tr>
<tr>
<td>a. Religious Affiliation and Minority Status ............................... 62</td>
</tr>
<tr>
<td>b. Substantive Nature of Dispute/Case Type ............................... 66</td>
</tr>
<tr>
<td>c. Judicial Characteristics .......................................................... 67</td>
</tr>
<tr>
<td>d. Other Variables ...................................................................... 69</td>
</tr>
<tr>
<td>Methods of Statistical Analysis ................................................. 72</td>
</tr>
<tr>
<td>Chapter 4. DESCRIPTIVE STATISTICS AND BIVARIATE ASSOCIATIONS .... 75</td>
</tr>
<tr>
<td>Introduction .............................................................................. 75</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Intercoder Reliability of Selected Measures ........................................... 56
Table 2: Denominational Protestant Categories ..................................................... 64
Table 3: Case Outcomes and Legal Tests in State Appellate Free Exercise Decisions ................................................................. 81
Table 4: Distribution of Claims by Substantive Nature of Dispute .......................... 88
Table 5: Logistic Regression Estimates of Bivariate Associations, 1997-2011 .......... 91
Table 6: Modeling the Likelihood of Successful Free Exercise Claims, 1997-2011 ..... 99
Table 7: Modeling the Likelihood of Successful Free Exercise Claims using Nominal Religious Classifications, 1997-2011 ........................................ 107
Table 8: Modeling the Likelihood of Successful Free Exercise Claims using Dichotomous Religious Classification, 1997-2011 ............................................ 109
Table 9: Modeling the Likelihood of Successful Free Exercise Claims, 2000-2011 .... 112
Table 10: Modeling the Likelihood of Balancing Test Use, 1997-2011 .................. 121
Table 11: Modeling the Likelihood of Balancing Test Use, 2000-2011 ................. 126
Table 12: Estimates from Structural Equation Model Predicting Balancing Test Use and Favorable Outcomes for Claimants, 1997-2011 ................................. 131
Table 13: Total, Direct and Indirect Effects of Predictors on Case Outcomes ......... 134
Table 14: Estimates from Trimmed Structural Equation Model Predicting Balancing Test Use and Favorable Outcomes for Claimants, 1997-2011 .................. 135
LIST OF FIGURES

Figure 1: Outcomes of State Appellate Free Exercise Claims, 1997-2011 ..................... 77

Figure 2: Distribution of Legal Tests Used to Evaluate Free Exercise Claims
After City of Boerne v. Flores, 1997-2011 ................................................................. 80

Figure 3: Comparison of Religious Groups By Share of Free Exercise Claims
and Representation in the U.S. Population, 1997-2011 ........................................... 83

Figure 4: Distribution of Free Exercise Claims Made By Denominational
Protestants .................................................................................................................... 84

Figure 5: Comparison of Successful Free Exercise Claims Across
Religious Groups ......................................................................................................... 86

Figure 6: Successful Free Exercise Claims by Case Type................................................. 89

Figure 7: Predicting Favorable Decisions by the Interaction of Religious
Group Size and Use of Judicial Retention Elections ............................................... 103
ACKNOWLEDGMENTS

This dissertation would not have been possible without the support, advice and patience of a great many people. I would especially like to express my gratitude to my advisor Roger Finke. His tireless support, guidance and motivation not only during the process of writing the dissertation but throughout my years at Penn State have been helpful beyond measure. I will be forever grateful for his mentorship and for his example, both as an academic and—more importantly—as a good man.

In addition, I would like to thank my doctoral committee members Duane Alwin, John McCarthy and Christopher Zorn. Their encouragement, advice and constructive comments made this a better project and made me a better scholar. I also wanted to thank Evelyn Bush, Allen Hertzke, Doyle McCarthy, Micki McGee, Emily Rosenbaum and Matthew Weinshenker for their support and time throughout the six years of graduate work that culminated in this dissertation. I would also be remiss if I didn’t thank Jessica Skocik, Gail Ulmer and all of the graduate students at the Association of Religion Data Archives for their help and understanding during the various stages of the dissertation process.

Special thanks to our families, especially my parents John and Susan Martin, my sister Juliann, and Sandi Barrale and Paul Solari, who gave me strength to see the dissertation through to its conclusion. Finally, none of this would have been possible without the support and understanding of my wife Danielle. Her patience and support over all the working weekends and late nights—especially during the final few months of writing, when we didn’t know whether our new child or the dissertation would be first on the scene—sustained me throughout this project and continues to do so. She amazes me every day.
During the writing of this dissertation, we lost two beloved grandparents,

Phyllis Schwartz and Albert Martin,

and gained one beloved son, Zachary,

who was born within one week of its completion.

The dissertation is dedicated to them.
CHAPTER 1: INTRODUCTION AND THEORETICAL BACKGROUND

Introduction

Religious believers across the globe have two sets of obligations: the obligation to follow the dictates of their creed and the obligation to obey the law of the state. Though these obligations are often in harmony, they come into opposition when conduct mandated under one set of obligations is proscribed under the other. Forms of religious practice can violate the secular legal order, just as compliance with the requirements of the state can violate dictates of conscience. American courts have tried to resolve many disputes of this type over issues as diverse as the days of the week believers can abstain from work,\(^1\) the ways their children are educated\(^2\) and the substances they can put in their bodies during religious worship.\(^3\) The federal Constitution and all state constitutions protect the right to practice one’s religion freely, but how durable are these protections in the face of laws that prohibit what believers feel they must do because of religion?

This dissertation explores the results of the judicial process Americans pursue when their religious and legal duties come into this sort of conflict. Outcomes of these judicial proceedings drive the ongoing struggle over how to define the expansiveness of rights to the free exercise of religion. In deciding which set of obligations trumps the other when the two are irremediably opposed, judges identify the boundary lines between what conscience demands and what the law requires. As a result, understanding variations in the success and failure of these kinds of free exercise claims in the courts is crucially important in allowing us to assess the status of this First

---

Amendment right in America today. What factors influence whether someone prevails in court when claiming that their free exercise rights have been abridged? And how much deference do judges give to constitutional protections of religious freedom when they are cited as justification to disobey the law?

The need for answers to these questions is especially pressing as they pertain to free exercise claims in the state courts. A flurry of legislative actions in recent years has led to different understandings what free exercise rights mean from state to state (Claborn 2011; Bridge, forthcoming). The dissertation contributes to our understanding of these questions through an analysis of quantitative data derived from state appellate legal opinions containing rulings on free exercise disputes. Through a systematic examination of court decisions from 1997 to 2011, this research deepens our understanding of how free exercise cases are decided in the state courts, explores whether some religious groups benefit more than others from the judicial process, and addresses whether the mechanisms at the heart of judicial and legislative efforts to broaden the scope of free exercise protections have had the impact they intend.

In Chapter 1 of the dissertation, I introduce the principal debate over the scope of free exercise rights. I also review how socio-legal theoretical frameworks in the legal positivist and realist traditions can assist us in understanding the sources of variation in both case outcomes and judicial understandings about how much protection free exercise guarantees offer. Next, I discuss the consequences of the judicial process through the lens of religious economies theory. Finally, I review the key research questions the dissertation explores, give a brief description of the research methods employed and data analyzed in testing hypotheses about these research questions, and offer an overview of the dissertation’s organization by chapter.
The Courts and the Meaning(s) of Free Exercise Rights

The debate over the proper interpretation of free exercise clauses at both the state and federal levels hinges which type of legal test judges should use in evaluating the facts of free exercise disputes. These two types of tests judges use in these cases are balancing tests and rational-basis tests. The tests differ considerably in what they require of those who allege violations of their rights to free exercise.

When judges make decisions in free exercise cases using a balancing test, they balance the severity of the burdens imposed by the government on claimants’ free exercise of religion against the government’s need to do so. They may find that the interest the government cites as its reason for imposing these burdens does not warrant their imposition, or that the government is justified in imposing these burdens because circumstances require them. The result is that courts may grant claimants exemptions from laws or government policies. Use of a balancing test acknowledges that citizens have obligations to their conscience as well as to the law, and that judges should require citizens to forsake their religious obligations only when necessary. The most common form of balancing test used in evaluating free exercise claims, as is explored greater depth below and in Chapter 2, is the “compelling interest test” articulated by the Supreme Court in its 1963 Sherbert v. Verner decision (374 U.S. 398).

The other form of test judges apply to the facts of free exercise disputes is the rational-basis test. This type of test does not permit exemptions from laws; rather, judges test whether the law or government act in question is rationally related to a secular interest of government. Application of this test is based on the proposition that constitutional free exercise clauses protect religious individuals and groups from religious discrimination, but that they are also required to follow the law if it is neutral regarding religion. The rational-basis test, as articulated
by the Supreme Court in its decisions in *Employment Division, Department of Human Resources of Oregon v. Smith* (494 U.S. 872 [1990]; referred to throughout as *Employment Division v. Smith*) and *O’Lone v. Estate of Shabazz* (482 U.S. 342 [1987]), represents a higher hurdle for free exercise claimants to scale. When judges apply this type of test, claimants must demonstrate not only that their free exercise has been burdened by the law, but that the law is discriminatory and is neither religiously neutral nor applicable to the population as a whole.4

Neither test dominates free exercise jurisprudence in state courts. As a result of legislative developments and court decisions over the last century, and particularly over the past two decades, the prevalent standard for evaluating free exercise claims varies considerably from state to state and from claim to claim. These developments are described in greater detail in Chapter 2, but a brief summary is necessary here in order to understand the questions the dissertation explores. First, the Supreme Court’s decision in *Employment Division v. Smith* (1990) required that a rational-basis test be used to evaluate most free exercise claims made under the First Amendment of the U.S. Constitution. This decision appeared to run counter to past precedent, as first articulated in *Sherbert v. Verner*, which required the application of the compelling interest test—a form of balancing test—in adjudicating these claims (McConnell 1990a:1110). The Court interprets the Constitution’s Fourteenth Amendment to mean that First Amendment freedoms are incorporated into state law5, and so the Supreme Court’s interpretation of the First Amendment in *Smith* was binding on state courts. However, Congress then recognized a backlash to this unpopular decision (Wood 1990:749-50) and enacted the Religious

---

4 An example of a successful outcome for claimants under this type of test is found in the Supreme Court’s 1993 decision in *Church of the Lukumi Babalu Aye v. City of Hialeah* (508 U.S. 520), in which the Court held that a city ordinance banning ritual animal sacrifice was not neutral with regard to religion. Once the city’s position failed the rational-basis test, the Court held that there was no compelling interest justifying the ordinance and declared it unconstitutional.

Freedom Restoration Act of 1993 (or RFRA; Public Law 103-141) to reinstate the compelling interest test as the dominant method of adjudicating free exercise cases at the federal and state levels. RFRA was in force for four years until, in its opinion in City of Boerne v. Flores (521 U.S. 507 [1997]), the Supreme Court held that RFRA could not apply to First Amendment claims in state courts. RFRA’s mandate was negated, and the rational-basis test articulated in Smith returned as the standard for adjudicating state free exercise claims based on the First Amendment.

The dispute over the meaning of free exercise rights in the states did not wind down with the Court’s decision in Flores, however, and it is the state cases decided after Flores that are the focus of this dissertation. Several developments after Flores further complicated the status of religious free exercise in the states, results of federal and state legislative action as well as state judicial intervention. For its part, Congress rolled back some of Flores’ control of state free exercise cases by enacting the Religious Land Use and Institutionalized Persons Act (2000, Public Law 106-274; also known as RLUIPA). This law had the same effect as RFRA in that it required judges to apply the compelling interest test when evaluating free exercise cases, but the law only controlled cases involving land use (i.e., zoning) disputes and the religious rights of the institutionalized. Other efforts regarded interpretations of the free exercise protections in state constitutions rather than the First Amendment’s Free Exercise Clause. About one-third of states enacted laws that have the same or similar effect as the federal RFRA, only regarding state-level free exercise provisions; these state RFRAs or “mini-RFRAs” require judges in those states to apply some form of balancing test when ruling on free exercise claims made on the basis of state constitutional protections. In addition, a number of state judiciaries interpret their constitutions’

6 The Supreme Court validated RLUIPA’s provisions concerning the rights of the institutionalized as constitutional in Cutter v. Wilkinson (544 U.S. 709 [2005]), while the Court has not dealt with the constitutionality of the law’s land use rights provisions as of Spring 2013.
free exercise protections as requiring the use of a balancing test as a matter of precedent; many of these judiciaries set this precedent after the Supreme Court’s decision to the contrary in *Smith*.7

The result of all this is that free exercise claimants today confront a patchwork legal opportunity structure at the state level, the shape of which varies by state and by the nature of the claim. For example, someone in Pennsylvania who objects on religious grounds to mandatory immunizations can make a free exercise claim under that state’s constitution, and because Pennsylvania has a state version of RFRA on its books, the claim will likely be judged using a balancing test. In contrast, the same claim in New Jersey will probably be evaluated using a rational-basis test, since the New Jersey Supreme Court held that this type of legal test should be used in these cases.8 Finally, claims involving the free exercise rights of prisoners in either state are controlled by federal legislation (RLUIPA) and will probably be judged using a balancing test for that reason. The net result is a great deal of heterogeneity in how judges evaluate religious free exercise claims in state courts across the country, which should have a dramatic impact on the conclusions they draw about the merits of these claims.

**Theoretical Frameworks: Legal Positivism and Legal Realism**

Understanding patterns in how these claims are decided, in terms of the legal tests judges apply as well as the conclusions they reach, is important for many reasons. First, examining case outcomes allows us to assess whether members of some faiths are unable to practice their religion with the same fullness that others can. Second, it allows us to verify that efforts to

---


expand the scope of free exercise rights, both by legislators and by precedent-setting judges, are aimed at the correct target: the type of legal test employed to evaluate free exercise claims. Third, a focus on state court opinions in the wake of *City of Boerne v. Flores* provides a window into the status of contemporary free exercise rights at the state level in ways that no other study I am aware of permits. Finally, the study provides an opportunity to assess competing arguments on the nature of judicial decision-making in disputes over civil liberties more generally.

On this last point, theoretical perspectives from legal theory and the scientific study of law guide the dissertation in its exploration of its focal research questions. These perspectives concern how laws should be understood and interpreted by both jurists and researchers alike. I focus on two distinct legal theoretical orientations that offer opposing conceptions on the nature and meaning of law in society: **legal positivism** and **legal realism**. These two orientations emerged in debates over legal philosophy, but also find expression in early work in the sociology of law.

Legal positivism regards the law as essentially a given, a social fact with weight and legitimacy by virtue of having emerged from a rational rule-making process. Kelsen ([1945] 2009) describes all law as positive law, as “its positivity lies in the fact that it is created and annulled by acts of human beings, thus being made independent of morality and similar norm systems” (114). In essence, laws are objective data to be ordered and interpreted (Hart 1961; see also Cotterrell 1992:8–9). Questions of the law’s appropriateness or morality are bracketed entirely; as summarized by Tamanaha (2007:23), “what law is and what law ought to be are separate questions.” Social scientific interpretations of the law with a positivist emphasis start from the premise that an understanding of the legal process should be guided by the “compulsive force of legal rules” (Mendelson 1963:593). Though they reject the criticism that such a view of
legal interpretation is “mechanical” or a deterministic interpretation of precedent or the letter of the law (Hart 1958:610-611; see also Dyzenhaus 2000:707), legal positivists regard the authority of rules as paramount in understanding the law and its interpretation. It is this approach that Segal and Spaeth (1993:33) refer to as the “legal” model of judicial decision-making.

Legal positivism shares an epistemological emphasis on value-free fact-finding with the formative sociological program of Auguste Comte, who envisioned sociology as the site of positive, empirical study of society separate from normative influences of other social spheres (see Coser 1977:12). Yet a number of the founding fathers of the sociology of law problematized the positive status of the law, arguing that the force of law is derived from social sources. Max Weber ([1922] 1978), for instance, examined the question of why laws are considered to be legitimate. As part of his larger project of examining the expansion of rationalization in society, Weber argued that modern law is validated internally through formal rationalization, that is, through its basis in general rules and procedures. It is also validated externally by staffs of specialists who ensure compliance with the law. Legal authority in modern states, then, derives from the generally applicable, codified (and hence knowable) character of the law.

Other founding legal sociologists cast doubt on the primacy of rules, which is at the core of the legal positivist orientation, by distinguishing between codified law and the law experienced in daily life. Leon Petrazycki, one of the most influential scholars in the Eastern European tradition of legal sociology, distinguished positive law from what he termed intuitive law, the mass of duties individuals acknowledge and are bound by that are not mediated by legal professionals ([1955] 2011; see also Treviño 2011). Eugen Ehrlich ([1913] 1962) made a similar distinction, conceiving of two types of law: juristic law, which emerges from lawmaking
institutions (especially the courts), and living law, the norms of conduct that dominate social life for the purposes of ensuring social cooperation. The latter has the ability to validate the former when they are in harmony. These early forays into understanding the social basis of law highlight a skepticism toward a legal positivist epistemology in which precedents or rules dominate how the law should be understood.

This skepticism receives greater expression in the tradition of legal realism, developed in legal theory and the scientific study of the law. Legal realism is a more demonstrably sociological approach to law, recognizing that interpretations of the law are colored by the experiences and preconceptions of its interpreters. Inspired by the early writings of Oliver Wendell Holmes, legal realism views precedents not as deterministic guidelines for future interpretations of the law, but as raw materials that influence these interpretations without controlling them (Deflem 2008:99). Karl Llewellyn (1931), one of the early proponents of this approach, agreed with positivist thought insofar as he believed questions of law’s ideal effects on society should be bracketed, but he emphasized that legal researchers needed to study its actual effects in practice. He put a particular emphasis on understanding the actions of legal professionals as the true expression of law in society, arguing that what officials “do about disputes is, to my mind, the law itself” ([1930] 1981:3).

A key tenet of legal realist theory is that understanding how these professionals arrive at legal conclusions is critical; according to this perspective, more formalistic interpretations of the law do not account sufficiently for the role of professionals’ agency in decision-making. A rich literature has emerged on judicial behavior and decision-making across the fields of legal studies and political science. Numerous factors may influence how judges interpret the law beyond their understanding of precedent, from personal attitudes to professional skills to strategic
considerations (Cotterrell 1992:219). Political scientists in the judicial behavioralist tradition fit squarely into the legal realist epistemology. The proposition that judicial values guide decision-making became a subject of analysis in the work of C. Herman Pritchett, who initiated the judicial behavioralist research program (Baum 2003:57). Pritchett’s work (1941, 1948) identified voting blocs in the Supreme Court based on judicial attitudes who commonly voted together; left, center and right voting blocs could be observed in the Franklin Roosevelt-era Supreme Court based on the frequency with which members of these blocs agreed in decisions (Pritchett 1948:55). This work relied primarily on comparing descriptive statistics such as agreement rates and raw numbers of votes.

Pritchett’s initial forays into the empirical analyses of judicial behavior based on judges’ attitudes inspired many researchers to explore this line of inquiry in greater detail. Glendon Schubert (1965:27) developed an attitudinal model of decision-making positing that court decisions were the result of judges’ attitudes toward policy matters and key dimensions of cases. Segal and Spaeth (1993:69), building off the work of Rohde and Spaeth (1976), modified Schubert’s model to conceive of values as an interrelated set of attitudes; the body of cases involving similar types of situations and objects constitute “issue areas,” and judges’ attitudes toward these issue areas should help explain their behavior. This conception of attitudes draws explicitly on Milton Rokeach’s definition of attitudes (see [1968] 1970:112) as enduring organizations of beliefs regarding objects or situations, which predispose people to respond in a preferential way toward these things. Scholars have subsequently attempted to measure these attitudes as well as identify how they affect decision-making.

Though some researchers have attempted to infer judges’ ideological values from textual materials such as newspaper editorials (Segal and Cover 1989:559), it is difficult to derive
precise measures of judicial values outside of considering judicial rulings. As Tate (1981:365) notes, deducing values from votes is possible, but it is not a useful strategy given that the strategy reverses the causal ordering most researchers aim to explore (i.e., that voting can be predicted as a function of ideology). This strategy is also not feasible for analysis of judges in the lower courts, given the paucity of non-opinion source materials for judges outside the Supreme Court.

One useful proxy for measuring judicial ideology is political affiliation. In a meta-analysis of relevant research from 1959 to 1998, Pinello (1999) found a substantial correlation between judges’ political affiliations and the hypothesized liberal or conservative direction of voting in each study (e.g., votes for defendants in criminal justice cases were coded as liberal, as were preferences for workers in cases involving labor disputes). Pinello observed a weighted mean correlation of .277 between political party and judicial ideology as inferred from voting patterns across 84 studies of American courts, and this effect size increased to .637 at the Supreme Court level. As would be expected based on Schubert’s attitudinal model, researchers have found powerful associations between judges’ ideologies and/or political affiliation and their votes on disputes over civil liberties. Segal and Cover (1989:561) observed a correlation of .80 between Supreme Court justices’ ideological values and their votes in these types of cases from 1953 to 1987. Likewise, Marshall and Ignagni (1994) found that the Court was less likely to support civil liberties claims as the proportion of Republican-appointed justices increased.

Another realist perspective to have emerged in recent years posits a rational-choice approach for explaining judicial decision-making. Judge Richard Posner (1986, 1993) has proposed that judges are rational actors who try to maximize the benefits they receive and limit costs. Cohen (1991:186) summarizes this proposition by arguing that “all else equal, judges
would prefer higher salaries, less crowded dockets, recognition within their legal and judicial circles and freedom to make decisions without outside interference.” Responding to Posner’s thesis, Schauer (2000) proposed a variety of motivations out of which self-interested judges might act. For instance, judges’ occupational ambitions should be examined as sources of behavior; judges may want to position themselves as best they can for promotion to seats on a higher court (Schauer 2000:631). Evidence suggests that judges rule in ways that maximize their likelihood of promotion. Cohen (1991:194) found that federal judges with greater promotion potential⁹ were more likely to vote to uphold federal sentencing guidelines, a politically popular decision. Likewise, Epstein, Landes and Posner (2013:363) found that judges with greater promotion potential were more likely to make politically safe decisions; they found that these judges were more likely to uphold convictions for ordinary criminal (“street crime”) offenses.

The two theoretical perspectives reviewed in this section – legal positivism and legal realism – offer competing insights that can assist in understanding factors leading to favorable decisions for claimants in religious free exercise cases. They may also help us understand the factors that influence judges in deciding the type of legal test to apply to these claims for much the same reasons. The hypotheses reviewed in Chapter 2 are derived from these two logics. In general, the hypotheses grounded in the positivist perspective flow from the premise that case outcomes and legal test selection are governed largely by pre-existing rules and precedent, and that observed variations in these outcomes are a function of neutral application of the law. For instance, as will be explored in the following section, religions groups that are more socially marginal should be less likely to prevail in their claims than other groups. This is due to the proposition that these groups’ religious practices are less likely to have been legitimated in the

⁹ Cohen (1991) measured promotion potential measured as a function of the number of district court positions available to judges from each state, the number of vacancies in relevant higher courts and the age of the oldest judge in these courts.
body of the law; thus, neutral application of the law as written and understood will result in a state of de facto legal inequality between religious groups. On the other hand, the legal realist perspective, largely understood in this dissertation within the framework of judicial behavioralism, suggests that studying judges’ attributes and the contexts in which they make their decisions can help us understand how the law is interpreted (Cotterrell 1992:218). The realist hypotheses tested are informed by the notion that judges’ attitudes toward civil liberties as well as their desire to maximize incentives influence their decision-making.

The Consequences of Decisions in Free Exercise Cases: Implications from Religious Economies Theory

The results of the hypothesis tests performed in this dissertation have broad implications, not just for the definition of free exercise rights, but for the ability of all religious groups to worship. In a country that protects the right to practice one’s religion freely, are members of some faiths not able to practice in the same fullness as others under the law? A body of social scientific literature examining the status of religious freedom in America has explored whether religious groups have an equal chance of receiving favorable decisions in their judicial free exercise claims (Way and Burt 1983; Brent 1999, 2003; Wybraniec and Finke 2001; Adamczyk, Wybraniec and Finke 2004; Sisk, Heise and Morriss 2004; Claborn 2011). Though the judgments of these studies are not unanimous, most suggest that members of smaller minority faiths are not only more likely to claim violations of their free exercise rights in court than members of more mainstream faiths, but are less likely to see their claims upheld. Insights from religious economies theory demonstrate why this divergence in outcomes is meaningful. This section provides a brief review of this theoretical perspective and analyzes why the question of
whether religious minorities are disadvantaged by interpretations of free exercise protections is worth exploring.

Religious economies theory emerged in response to the older paradigm of how religion should be studied sociologically. As summarized by Warner (1993), work in the older paradigm viewed monopoly religions as the model form of religious organization; as such, it was dominated by concerns over monopoly maintenance and groups’ ability to remain plausible in the face of competing cosmological explanations (e.g., Berger 1967). The newer perspective conceives of a society’s religious sphere as a religious economy, which is populated by all present and potential religious adherents, religious organizations seeking to retain and/or attract these adherents, and the religious cultures promoted by these organizations (Finke and Stark 1988, 2005; Stark and Finke 2000:193).

A fundamental proposition in this perspective is that potential religious adherents are self-interested actors who try to maximize benefits and minimize costs. Religious organizations offer benefits that other kinds of organizations cannot; these are what Stark and Bainbridge (1985) term “compensators,” which are often of a supernatural character and provide deferred rewards (e.g., access to heaven, salvation or prosperity) in exchange for certain behaviors. Theories about how such organizations enjoy success have evolved. Iannaccone (1994), revisiting early work by Kelley (1972) on this subject, argued that organizations that demand a significant level of commitment and sacrifice yield a more satisfying religious product than those that do not. This initial cost selects out less-committed members, minimizes “free riders” and populates the organization with enthusiastic members. Building on Johnson’s (1963) distinction between churches and sects, Stark and Finke (2000) revised this model somewhat by arguing that potential adherents vary in terms of how much they want to belong to religious organizations.
operating at different levels of tension with the greater society (i.e., the degree to which organizations function in opposition to mainstream society). The distribution of demand for different degrees of tension is assumed to be normally distributed. According to their model, religious organizations enjoy the greatest success when they function at a moderate yet pronounced level of tension with society; however, due to pressures from organizational elites to lower this tension level, organizations rarely maintain this level of tension over a long period of time.

As in most economies, regulatory bodies structure the operation of the market for adherents, setting the rules by which religious organizations function and by which adherents practice their religion. Government is the primary regulatory actor that establishes and maintains the rules of religious economies. It may do so through restrictions on how religious bodies operate, which may involve anything from limitations on certain behaviors (e.g., the wearing of religious garb) to outright bans on religious groups. It also does so through shows of favoritism to some groups; such favoritism can take many forms, from financial subsidies to decrees that only one religious faith is permitted to operate under the law. The primary effects of religious regulation are to alter the structure of opportunities and incentives for religious organizations and to limit the variety of membership options for potential adherents (Finke 1990).

Three decades of research employing this perspective highlight the implications of state regulation on the exercise of religion (Finke 1990; Iannaccone 1991; Chaves and Cann 1992; Stark and Iannaccone 1994; Finke 1997; Stark and Finke 2000; Fox and Tabory 2008). Several studies have shown a positive link between how freely religious organizations compete for adherents and the strength of religious participation in those organizations (Finke and Stark 1988; Hamberg and Pettersson 1994; Olds 1994; however, see Land, Deane and Blau 1991 and
Montgomery 2003 for a contrary view). Moreover, there is evidence that restrictions or limitations on religious practice depress activity in the religious market, or at least drive it underground. Fox and Tabory (2008) found that more robust government restriction on religion internationally is associated with lower levels of service attendance. Conversely, Gill’s (1999) study of countries in Latin America greater degrees of religious pluralism where levels of religious regulation by the state were lower.

Finke and Stark (2005:11-12) have characterized the religious economy in the United States as largely deregulated, which has allowed newer religious organizations to thrive and has given rise to a tremendous level of religious innovation. Efforts to quantify the degree of regulation in the American religious economy compared to other countries echo this assessment. In his cross-national survey of government involvement in religion, Fox (2008) observed that the United States was the only one of the 175 nations he studied to earn the highest scores on all of his measures of church-state separation. Although it did not list the United States as one of the very lowest-scoring world nations on their measure, a report from the Pew Research Center’s Forum on Religion & Public Life (2011:91) noted similarly low levels of government restrictions on religion in 2008 and 2009.¹⁰

Laws in America cannot purposefully or overtly discriminate against individuals due to their religion, nor can they intentionally restrict religious practice. This is true regardless of the type of legal test judges apply in evaluating disputes over constitutional free exercise rights. However, inadvertent restrictions on religious practice do exist. The vast majority of free exercise disputes heard in the courts concern laws or government policies that allegedly burden people’s or organizations’ free exercise rights in the course of the government pursuing some

¹⁰ The United States earned scores of 1.6 on the Government Restrictions Index (GRI) used in the Pew Forum’s report during the periods ending in both mid-2008 and mid-2009; potential scores ranged from 0 to 10, with a score of 10 indicating the maximum degree of restriction.
broader secular aim, such as public safety or the protection of children. Religious economies theory postulates that this is particularly worrisome for members of religious minority groups, who are disproportionately likely to be burdened by these sorts of incidental restrictions (Wybraniec and Finke 2001:430-431). Judges’ choice of the legal test used in adjudicating claims should be of particular concern to religious minorities, for as Donald Black (1976:114) has argued, outsiders are less likely to enjoy the benefits of the law than those who enjoy greater acceptance and respectability. Those with less power to influence the process of lawmaking will likely find themselves burdened more heavily, albeit incidentally, by laws that do not have the expressed goal of restricting religious practice. Incidental burdens on religious exercise, which should be more frequent for members of religious minority groups, are not remediable when claims are judged under a rational-basis test, but they may be under a balancing test.

An analysis of free exercise case outcomes, then, reveals the terms of debate over the scope of constitutionally protected religious practice. While no legislative body in America may criminalize religious behavior for its own sake, it may enact laws that effectively limit religious behavior in order to further a secular government interest—and due largely to their position out of the center of religious and political life in the United States, religious minorities’ practice may be disproportionately restricted. It is for this reason that understanding free exercise claims in the courts is important, because judiciaries can offer protections for minority rights that other branches of government may be unable or unwilling to provide (Richardson 2006; Finke and Martin 2011). Autonomous, independent judiciaries allow members of minority faiths the opportunity for equal time to plead their cases and receive a fair hearing; likewise, these bodies are better positioned to withstand backlash from societal majorities if they uphold the rights of unpopular groups. However, previous research demonstrating the difficulty minority faiths
experience in pursuing free exercise claims in the courts raises troubling questions about the state of religious freedom in America. A disparity in outcomes between minority faiths and more mainstream faiths would result in a de facto favoritism toward majority religions, which would have profound effects on the American religious landscape. It would erect barriers to minority religious groups’ ability to worship and to propagate their faiths, and would limit the range of religious choices available to potential adherents.

Summary and Organization of Dissertation

This dissertation is driven by two principal research questions, which are:

1. What factors predict whether a free exercise claim will be upheld or rejected in the state appellate courts?

2. What factors predict whether judges will rule on these claims using a rational-basis test or a balancing test?

The first question has been explored at a national level in a number of previous studies (e.g., Way and Burt 1983; Wybraniec and Finke 2001; Sisk, Heise and Morriss 2004), but the present study expands upon these works using contemporary data, additional predictors and a specific focus on the state appellate courts. The only exclusive study of state courts conducted recently (Claborn 2011) examined “religious freedom” cases more broadly; the dissertation has a narrow focus on cases containing claims made under state or federal constitutional protections of religious free exercise. Chapter 2 of the dissertation explores findings from previous research on this topic in greater detail. It also provides an in-depth historical background on free exercise litigation in the United States, with particular emphases on the key milestones marking how the definition of the free exercise of religion has evolved and the significance this evolution for
state-level litigation. Finally, Chapter 2 introduces the hypotheses tested in the dissertation, which proceed from the legal positivist and realist perspectives on judicial decision-making. One of the dissertation’s most important contributions is to account for the use of the two classes of legal tests that judges employ in evaluating free exercise claims, tests which pose different degrees of difficulty for claimants depending upon the site and substance of the claim. The application of one test over another should yield stark differences in case outcomes: one type of appears to leave little space for claimants to prevail on free exercise grounds, while the other appears to allow for far greater accommodation of religious exercise.

Hypotheses are further categorized by which of the two research questions they address. In investigating my second research question, the dissertation treats the application of rational-basis or balancing judicial tests as problematic and will investigate why one test or the other is applied across cases. While earlier research has acknowledged the importance of different legal standards for influencing how cases are resolved (Wybraniec and Finke 2001; Adamczyk, Wybraniec and Finke 2004), the proposed study will be the first to this author’s knowledge to model the likelihood that a particular standard will be adopted in a given free exercise case. I employ structural equation modeling techniques to analyze what predicts the use of different legal tests and different case outcomes simultaneously. An empirical investigation of the judicial protections afforded to those who claim that their rights to practice their religion have been abridged will contribute to understandings of both (a) the state of the free exercise of religion today and (b) the factors that benefit or impede the religious as they try to resolve their disputes in court.

Chapter 3 addresses my methodological strategies for testing the hypotheses presented at the end of Chapter 2, focusing on the use of binary logistic regression models in predicting the
outcomes of interest as well as on structural equation modeling techniques employed to predict these outcomes simultaneously. Chapter 3 also describes the sources of data analyzed in the dissertation, which are comprised of data from secondary sources and from my own content analysis of court decisions. Data from the latter source, which include the two outcome variables, are derived from systematic coding of all published court decisions from state appellate courts from mid-June 1997 through December 31, 2011, located by searching for cases in the LoislawConnect primary source database. In all, I coded some 453 decisions for this analysis, of which 225 were decided using either a rational-basis or balancing test. This dataset, as well as other secondary datasets important for my analysis, is discussed in detail in Chapter 3.

In Chapter 4, I present key descriptive findings from the data and discuss basic bivariate associations. This chapter characterizes patterns of outcomes of state free exercise cases, both on a univariate basis and in conjunction with other measures. It also addresses the distribution of cases decided under the two forms of legal tests, the types of claims heard by courts and other univariate distributions. Further, Chapter 4 describes patterns of claims made from across the American religious landscape, examining (a) whether religious groups bring claims in proportion to their numerical size in the population and (b) whether members of minority faiths enjoy similar success rates in pursuing their claims as members of more mainstream faiths.

Chapter 5 is one of two chapters that employ multivariate analysis in answering the dissertation’s two primary research questions. This chapter takes up the first research question, analyzing the factors that predict whether free exercise claims will be decided in claimants’ favor in the state appellate courts. The binary logistic models presented in Chapter 5 evaluate competing hypotheses on this issue, examining the predictive power of legal test use, claimant and case characteristics (namely religious minority status and the substantive basis of the case),
judge characteristics and the context in which cases are decided (e.g., the religious composition of the communities where claims are heard). Chapter 6 turns to an analysis of the second research question, examining the factors that influence the type of legal test judges use in evaluating free exercise claims. The chapter reports findings from binary logistic regression analyses that attempt to answer this question directly as well as results of structural equation models that address each of the two research questions simultaneously. Chapter 7 concludes the study by reflecting on the dissertation’s findings and discussing the analysis’ implications for the state of religious free exercise in America, examining some of the study’s limitations and discussing additional avenues for further investigation.
CHAPTER 2: FREE EXERCISE RIGHTS – HISTORY AND HYPOTHESES

The previous chapter introduced the two principal research questions I attempt to answer in this dissertation: what are the factors that predict (1) whether a free exercise claim will be upheld or rejected in the state courts and (2) whether judges will use a rational-basis test or a balancing test to make their decisions. Chapter 2 provides additional background that is necessary in order to explore these questions. First, I review the past and present legal opportunity structure free exercise claimants confront in the state courts, placing particular emphasis on the historical emergence of the two types of legal tests employed to evaluate free exercise claims today. Next, in order to determine what factors have been associated with particular patterns of free exercise case outcomes in the past, I review findings from previous studies on this line of inquiry. Finally, I present formal hypothesis statements about the factors I expect to be linked to variations in case outcomes and legal test use; these hypotheses are based upon the legal positivist and legal realist theoretical frameworks presented in Chapter 1.

The Courts and the Free Exercise of Religion: A Primer

The meaning and scope of constitutional free exercise protections was not notably controversial until the mid-to-late 20th century. Until recently, courts varied little in how they understood the definition of free exercise rights vis-à-vis citizens’ obligations under the law. Yet the Supreme Court’s apparent divergence in its 1990 Employment Division v. Smith decision from its previous free exercise jurisprudence (McConnell 1990b) and the subsequent responses of lawmakers and state courts have thrown two contrasting interpretations of free exercise rights in America into relief. Some state courts, following the Supreme Court’s reasoning in the Smith
decision, hold that these rights do not give citizens license to disobey generally applicable laws; only when laws intentionally discriminate against religious groups do these courts support claims that religious obligations supersede legal obligations. Other state courts, in keeping with the mid-20th century conception of free exercise rights, hold that exemptions to laws or government policies are permissible even if those laws or policies impose merely incidental burdens on free religious practice. To understand how this divergence manifests itself in the present-day state-level legal opportunity structure, it is necessary to review the evolution of how free exercise rights have been understood in America. Such a review must focus predominantly on decisions at the federal level and namely on Supreme Court decisions, since the states had largely followed the Supreme Court’s lead on how to interpret even state constitutional free exercise provisions until recently; however, I incorporate key developments at the state level into this review where appropriate. This approach allows for a thorough accounting of how the contemporary opportunity structure free exercise claimants face in the state courts developed over time.

I. Early Free Exercise Jurisprudence: 1813-1963

McConnell (1990b) identifies the earliest state court decision of which we have a thorough record, an 1813 matter involving a Catholic priest who was subpoenaed in New York to appear before a grand jury. A man had confessed an act of theft to the priest and had given him the stolen goods to return to the victim; upon receiving his goods again, the victim informed the authorities of what had happened. The priest refused to testify when subpoenaed, saying he would rather die than disclose the subject of a sacramental confession. The court agreed with the priest’s contention that he should not be required to testify, ruling in People v. Philips (cited in
McConnell 1990b) that clergy were exempt from revealing the subjects of confessions made to them in their ministerial duties.

The Supreme Court, however, did not hear a case addressing the scope of the First Amendment’s Free Exercise Clause until 1878. In its decision in Reynolds v. United States (98 U.S. 145 [1878]), the Court established the precedent that while government cannot regulate religious beliefs, it may restrict action based on those beliefs when pursuing some broader interest. The Reynolds case concerned whether or not a Mormon man with two wives could be punished for violating a federal ban on polygamy, even though his religious convictions mandated that he take multiple wives as a matter of religious obligation. The Court held that the ban on polygamy was valid in the interest of upholding a moral and democratic society; the First Amendment, it held, did not give anyone the right to disobey the law of the land (Flowers 2005).

Yet the precedent established in Reynolds was not to be the guiding standard for evaluating free exercise claims through most of the 20th century. The government’s ability to restrict religious behavior was narrowed considerably in 1940, when the Court held in its Cantwell v. Connecticut (310 U.S. 296) decision that religious practices could only be restricted by government if they constituted a “clear and present danger” to the public order (Flowers 2005:25). Restrictions on religious practice were now only permissible if they were deemed necessary to ensure the public safety. The decision established the precedent that courts needed to balance religious obligations with citizens’ obligations under the law; in other words, Cantwell mandated that judges use some form of balancing test in evaluating free exercise claims. The case concerned a New Haven, Connecticut ban on selling religious materials in public without prior approval from local officials. A family of Jehovah’s Witnesses was indicted for violating this ban, but the Court held that the city’s ban violated the First Amendment. Moreover, the
Court concluded that the Due Process Clause of the Fourteenth Amendment meant that First Amendment freedoms applied in the states as well, and that state legislatures could not impose restrictions on freedoms guaranteed by the Constitution.¹¹

II. The Rise of the Compelling Interest Test: 1963-1990

The Court’s incorporation of First Amendment rights into state law in Cantwell v. Connecticut gave state courts the ability to decide cases involving federal free exercise protections, and it also deepened the volume of free exercise cases heard by federal courts. Indeed, most of the landmark decisions on free exercise rights involved disputes with individual states that were ruled upon by the Supreme Court. This expansion of the Court’s jurisdiction allowed it to hear a dispute involving the state of South Carolina’s unemployment compensation policies in 1963; its decision in this case established the dominant standard for evaluating free exercise claims over the next few decades. In Sherbert v. Verner (374 U.S. 398 [1963]), the Court heard the case of Adell Sherbert, a Seventh-day Adventist who lost her job because of her refusal to work on the Sabbath, which Seventh-day Adventists observe on Saturdays. Unable to find another job that permitted her to stay home from work on Saturdays, she applied for unemployment compensation from the state of South Carolina. The state denied her request on the grounds that she was able to work and was choosing not to do so; she responded by suing the state on free exercise grounds, and her case ultimately reached the Supreme Court.

¹¹ Sections 1 and 5 of the Fourteenth Amendment informed the Court’s interpretation. The relevant portion of Section 1 reads that “No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.” This passage is incorporated in state law in Section 5, which reads that “The Congress shall have power to enforce, by appropriate legislation, the provisions of this article.”
The Court ruled in favor of Ms. Sherbert, and in doing so, it developed the three-part test for ruling on free exercise cases which became known as the “Sherbert test” or the “compelling interest test.” Under this test, a claimant must first establish that a government actor, through law or enforcement of policy, has imposed a substantial burden on their sincerely-held religious beliefs. If this condition is met, the government actor has to demonstrate that the burden was justified in order to further a compelling state interest, or an interest of paramount importance to the government; otherwise, the claimant will prevail. Finally, if the government actor can prove this, they must further demonstrate that the method used in pursuing the compelling interest imposes the lowest possible level of restrictions on the claimant’s religious practice; in other words, the method was the least restrictive means of serving the governmental interest. Free exercise claimants prevail under this form of balancing test if they meet the first requirement and the government does not meet the second and/or third requirement. However, this type of test, also referred to as strict scrutiny, still permits the government to impose restrictions on religious behavior. Even if claimants demonstrate that the government placed substantial burdens on their ability to practice their religion, courts may rule that the burden was justified if the government actor meets the second and third requirements of the test (i.e., burdens are justified by a compelling government interest, and the government employed the least restrictive possible method of furthering that interest) (Flowers 2005:31-3).

The compelling interest test established in Sherbert v. Verner was upheld and reinforced by the Court in its 1972 Wisconsin v. Yoder (406 U.S. 205) decision. In that case, a number of Amish citizens asked for permission to withdraw their children from the public schools after eighth grade, an impermissible request under the state of Wisconsin’s compulsory education laws. Applying the Sherbert test to the facts of the case, the Burger Court held that the state did
not have a compelling interest that would justify compulsory school attendance in the face of religious objections, and it decided in favor of the free exercise claimants (Flowers 2005:34). The Court’s decisions in Sherbert and Yoder continue to be cited in state courts as guides for how free exercise claims based on state constitutional protections should be evaluated. However, claims based on the First Amendment in the state and federal courts could only be ruled upon using the Sherbert test up to 1990, when the Supreme Court issued a decision that appeared to reverse its jurisprudence dating from Cantwell v Connecticut.

III. The Contest Over the Meaning of Religious Free Exercise: 1990-1997

In 1990, the Supreme Court surprised some observers (McConnell 1990a; Choper 1992) by explicitly affirming a rational-basis test for settling most free exercise claims, which narrowed the applicability of the Sherbert compelling interest test dramatically. In Employment Division, Department of Human Resources of Oregon v. Smith (494 U.S. 872 [1990]), two substance abuse counselors were relieved of their duties for taking part in a peyote ritual, a component of the Native American Church’s worship services. Their subsequent unemployment compensation claims were rejected by the state on the grounds that they had been fired for violating the state of Oregon’s criminal ban on peyote use. As in Sherbert, the fired drug counselors responded by suing the state, and their case was granted certiorari by the Supreme Court (Marin 1991). The Court rejected the counselors’ free exercise claim without making any attempt to balance the counselors’ free exercise rights against Oregon’s state interests. They instead based their decision on the principle that the ban on peyote was a law that was neutral in its application toward religion (i.e., it did not have a discriminatory purpose) and was generally applicable to
all. Since the Court held that exemptions to laws meeting these criteria are not permitted by
under the U.S. Constitution, the counselors lost their case (Laycock 1994:887-888)\textsuperscript{12}.

Justice Scalia’s majority opinion attempted to square this conception of free exercise
rights with the Court’s earlier rulings by arguing that those cases constituted exceptions to the
Court’s 1878 Reynolds standard. According to Scalia, earlier cases evaluated under the
compelling interest test had involved either hybrid combinations of free exercise claims and
types of other rights claims (e.g., the rights of parents to be responsible for their children’s
education in the Yoder case) or individualized assessments made by the government (e.g.,
whether the state should have granted unemployment benefits in the Sherbert case). These
allegedly exceptional cases remained subject to the form of balancing test articulated in the
Sherbert decision (Forren 2006). But as legal scholar Michael McConnell (1990a:1153)
observed, Scalia’s opinion effectively replaced the prior standard for adjudicating free exercise
claims with “a bare requirement of formal neutrality. Religious exercise is no longer to be
treated as a preferred freedom; so long as it is treated no worse than commercial or other secular
activity, religion can ask no more.”

The decision provoked a hostile reaction from many quarters. A petition was delivered to
the Court in opposition to the decision and was signed by a kaleidoscopic array of religious
bodies (from the Unitarian-Universalist Association to the Baptist Joint Committee on Public
Affairs), constitutional scholars and think-tanks (Wood 1990:749-50). The result of the outcry to
the Smith decision was the nearly unanimous passage of the Religious Freedom Restoration Act
(RFRA) in 1993 in both houses of Congress. Signed into law by President Clinton, RFRA
mandated that the Sherbert test be used by the courts to adjudicate free exercise cases (Laycock

\textsuperscript{12} Nevertheless, within two years of the Court’s decision, the state of Oregon reversed course and exempted the
sacramental use of peyote from criminal bans on the substance for everyone except the incarcerated (Laycock
and Thomas 1994). RFRA was in force through the mid-1990s until its constitutionality came into question when the Supreme Court heard arguments in the *City of Boerne v. Flores* (521 U.S. 507 [1997]) case.

**IV. City of Boerne v. Flores and a Patchwork of Meanings: 1997 to 2013**

The Court responded to the passage of RFRA in 1997 in *City of Boerne v. Flores* (521 U.S. 507). In his majority opinion, Justice Kennedy found that RFRA as it applied to the state courts was unconstitutional; Congress had overstepped its bounds in instructing state courts how to interpret the Constitution. *Flores* concerned a dispute over a Catholic church that wanted to expand its physical plant but was prevented from doing so by local zoning regulations (Waltman 2011). The *Flores* decision reaffirmed the use of rational-basis test articulated in *Smith* for adjudicating free exercise claims in the state courts. However, it left the appropriate test for adjudicating federal free exercise cases ambiguous., and nearly a decade would pass before the Supreme Court declared that RFRA was constitutional as it applied to cases in federal courts (*Gonzales v. O Centro Espirita Beneficente Uniao do Vegetal*, 546 U.S. 418 [2006]).

As introduced in Chapter 1, there have been a number of efforts taken to restore the compelling interest test articulated in *Sherbert* to application in the state courts since 1997. First, in reaction to the decision that RFRA did not control state judicial interpretations of federal free exercise protections, many states have enacted legislation or amended their constitutions to codify RFRA’s principles into state law (Nugent 2008:1052). In general, these so-called

---

13 As of May 2013, sixteen states have enacted state RFRA legislation, while a seventeenth state (Alabama) amended its constitution with the language of the federal RFRA. In addition to the “mini-RFRA” legislation enacted in Connecticut in 1993, the states of Arizona, Florida, Idaho, Illinois, Missouri, New Mexico, Oklahoma, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas and Virginia all enacted mini-RFRAs following the Court’s decision in *City of Boerne v. Flores*. Louisiana (http://www.legis.la.gov/legis/BillInfo.aspx?s=10RS&b=SB606&sbi=y) and Kentucky (Musgrave 2013) enacted legislation of this kind after Lund’s review of state RFRAs. Finally, Lund notes (2010:477) Utah adopted a version
“mini-RFRAs” require judges to use the compelling interest test in adjudicating free exercise claims based on state constitutional protections. They do so by stating that the free exercise of religion in a given state may not be substantially burdened by facially neutral and generally applicable laws unless it is done in the name of a compelling government interest, which must be pursued using the least restrictive possible means.

In addition, Congress passed the Religious Land Use and Institutionalized Persons Act (or RLUIPA) in 2000, a measure compelling judges at all levels to use the compelling interest test in free exercise cases involving (a) prisoners’ rights and (b) land use disputes (Jelen 2010:84-85). Prior to RLUIPA, prisoners’ free exercise claims were subject to a rational-basis standard controlled by precedents other than Employment Division v. Smith. According to the precedents set in the Court’s 1987 Turner v. Safley (482 U.S. 78) and O’Lone v. Estate of Shabazz (482 U.S. 342) decisions, prisoners’ free exercise claims could only be upheld if prison officials were unable to justify their restrictions on inmates’ religious liberties as being reasonably related to legitimate penological interests. This rational-basis test is analogous to the one articulated in Smith in that discrimination on a religious basis was not permitted under constitutional free exercise protections, but these protections did not allow prisoners to be exempt from policies that apply to others on religious grounds. RLUIPA was an attempt to supplant the rational-basis test mandated by the Turner and O’Lone precedents with the compelling interest test articulated both in Sherbert v. Verner and in the Religious Freedom Restoration Act (RFRA).

Yet RLUIPA was far less ambitious an effort than RFRA due to its narrow focus on prisoners’ rights and land use free exercise disputes (excluding all other types of cases). The

of RFRA that only pertains to land use disputes; I do not count the Utah law as a state RFRA because of its limited scope.
piecemeal nature of the federal response to the Court’s *Flores* decision was the result of resistance by civil rights groups to a full-scale reworking of the original RFRA. Pushback was especially strong from groups that were concerned about the meaning of such a law for gays and lesbians, who feared that the law could lead to legalized discrimination by those alleging religious objections to homosexuality (Waltman 2011).

Finally, it should be emphasized that some state judiciaries interpret the free exercise provisions in their state constitutions to require the use of some form of balancing test, be it the compelling interest test or something less protective, without any state or federal legislation compelling them to do so. Every state has its own constitution, each of which contains a clause or provision protecting the free exercise of religion. State courts are free to interpret their own constitutions as they see fit, and some state supreme courts (Alaska\(^\text{14}\), Massachusetts\(^\text{15}\), Washington\(^\text{16}\) and others) even after the *Smith* decision decided that a form of balancing test should be used when evaluating free exercise claims based on state constitutional protections. I have come across a substantial number of decisions in the course of conducting this study in which claimants invoked the protections of both the state and federal free exercise clauses; in many of these cases, claimants’ First Amendment claims were adjudicated using the rational-basis test under *Smith*, while their state constitutional claims were adjudicated using a balancing test.

Thus, present-day free exercise cases are typically subject to one of two classes of legal tests. Under one class—rational-basis tests, such as the one articulated in the *Smith* decision—no exemptions to laws can be made on free exercise grounds if the laws are facially neutral toward religion and are generally applicable. Under the other—balancing tests, such as the one

\(^{14}\) *Swanner v. Anchorage Equal Rights Com’n*, 874 P.2d 274 (1994)
\(^{15}\) *Attorney General v. Desilets*, 636 N.E.2d 233 (1994)
\(^{16}\) *First Covenant Church v. City of Seattle*, 840 P.2d 174 (Wash. 1992)
articulated in the Sherbert decision—religious exemptions are possible if the burdens these laws impose on religious exercise outweigh the government’s interest in imposing them.

**Review of Free Exercise Research**

Given the opportunity structure claimants face in the courts, whose claims are upheld and whose are rejected? A number of sociologists, political scientists and legal scholars have tried to answer this question in various contexts over the last thirty years. I briefly review findings from these studies below. These studies inform my hypotheses and reveal unanswered questions as well as areas where more contemporary data and a narrow focus on free exercise disputes can add to our understanding of how free exercise cases are decided. I present my hypotheses following this review of the literature.

Way and Burt (1983) conducted the first comprehensive study of free exercise case outcomes in an attempt to study how the judiciary can legitimate marginal religious groups. Countering Pfeffer’s (1974) thesis that such groups are socially legitimated only when they make their religious doctrine more acceptable to broader society or when social norms change, Way and Burt argued that legitimation occurs through judicial acknowledgment that these groups’ practices are legally permissible. This thesis is consistent with pre-Smith jurisprudence at all levels, when balancing tests were used extensively and almost exclusively throughout the American courts in evaluating free exercise disputes. Way and Burt compared outcomes from all published state and federal free exercise decisions in two legal periods: a period prior to the Supreme Court’s Sherbert v. Verner decision (1946 through 1956) and a period following this decision (1970 through 1980). They found sharp increases with time in both the number of claims litigated and the proportion of claims that were upheld; while claimants received
favorable decisions in approximately 12 percent of cases in the earlier period, they received favorable decisions in 37 percent of cases in the post-*Sherbert* era.

These researchers were particularly interested in differences in outcomes between marginal and more socially accepted religious groups, which they examined on a bivariate basis. This yielded the surprising finding that marginal religious groups had greater rates of success in the courts from 1970 to 1980 than did more mainstream groups (Way and Burt 1983:657). This result appears to be a function of the types of claims these marginal groups raised; Way and Burt found that these groups were more likely to be involved in prisoners’ rights claims and particularly claims over the right to proselytize. Claims of the latter type were made almost exclusively by members of marginal religious groups, and perhaps due to the precedent set in *Cantwell v. Connecticut* (310 U.S. 296 [1940]), claimants won these types of cases overwhelmingly.\(^\text{17}\) Way and Burt’s finding regarding marginal group successes is unique among all studies of free exercise cases that examined rates of success across religious groups, perhaps because it is the only study to have analyzed decisions made entirely before the Court’s 1990 *Employment Division v. Smith* decision.

James Brent explored patterns in free exercise outcomes in the U.S. Courts of Appeals on two occasions, first in the wake of the Religious Freedom Restoration Act’s passage (1999) and again after the Court’s decision in *City of Boerne v. Flores* (2003). His earlier study is the first of its kind that I am aware of to employ multivariate statistical methods to control for competing explanations of free exercise outcomes. Employing a principal-agent theoretical framework to understand how lower courts act as agents working on behalf of higher courts, Brent investigated

\(^{17}\) It is worth noting that a far lower percentage of the free exercise decisions I read and coded in my analysis involved the right to proselytize or express one’s faith publicly. It is possible that the government was less likely after 1997 to litigate disputes over this issue given that they were very unlikely to prevail in these cases during the period before 1980.
how the Courts of Appeals behaved when faced with competing dictates from multiple principals (the Supreme Court and Congress). In both studies, he found that the Courts of Appeals were more likely to follow the lead of the Supreme Court on free exercise matters than Congress; they were significantly less likely to rule in favor of claimants after the Smith decision than before it, but were no more likely to rule in favor of claimants after RFRA was passed than in the pre-Smith period (Brent 1999:254). The latter comparison is perhaps misleading, since courts in the period following RFRA were mandated to use the same legal test in evaluating claims as they were in the pre-Smith period; as such, there is little reason to suspect that the likelihood of claimant success should be substantially different between these two periods. Brent also observed that courts were less likely to uphold free exercise claims in the period following the Court’s 1997 City of Boerne v. Flores decision than in the period after RFRA was in force (Brent 2003:566).

Brent’s work is also useful in that it tests additional explanations for free exercise successes. For instance, like Way and Burt (1983), he evaluated differences in outcomes based on claimants’ religious affiliation using a dichotomous indicator using binary logistic regression, and found the opposite result; members of mainstream religious faiths (defined by Brent as Catholic and “major” Protestant groups, which he considers “Presbyterians, Baptists, Lutherans, Episcopalians, etc. [259]) were significantly more likely in both studies to receive favorable decisions from the courts than non-mainstream faiths. Additionally, he examined whether claims involving prisoners’ free exercise rights were less likely to be upheld than other types of claims, given that these claims were not subject to review under the compelling interest test prior to 2000. However, he found no significant differences in outcomes between prisoners’ claims and other types of claims in either study (Brent 1999:254; 2003:566). Finally, although Brent
observed a significant, positive correlation between successful outcomes for claimants and Democratic majorities on the courts hearing claims (Brent 1999:251), he found no differences in likelihoods of success for claimants in Democratic- and Republican-majority courts in either study.

Relying primarily on data generated through content analysis of state and federal free exercise decisions by Wybraniec (1998), studies by Wybraniec and Finke (2001) and Adamczyk, Wybraniec and Finke (2004) explored differences in outcomes across three time periods: the decade prior to the _Smith_ decision (1981 to 1990), the period between _Smith_ and the passage of RFRA (1990 to 1993) and the period following RFRA’s passage (1993). Consistent with Brent (1999), both studies observed that free exercise cases heard from 1990 to 1993 yielded a smaller proportion of favorable decisions for claimants than in the period prior to the _Smith_ decision (Wybraniec and Finke 2001:436; Adamczyk et al. 2004:250). Moreover, Wybraniec and Finke (2001:437) found that favorable decisions for free exercise claimants were significantly more likely in the period after the passage of RFRA than in the 1990-1993 period when the rational-basis test articulated in _Smith_ was the dominant legal standard for evaluating claims.

The two studies based on Wybraniec’s (1998) dataset introduced the concept that the outcomes of free exercise disputes in the courts have major implications for the state of the American religious economy. These researchers examined differences not only in the outcomes of cases between minority and mainstream religious groups, but in the number of claims raised by different types of religious groups. Wybraniec and Finke (2001:434) observed that members of Protestant sects as well as a number of minority religious faiths (Muslims, members of Native American religions, and especially members of new religious movements) were disproportionally represented in the universe of free exercise cases heard by state and federal
courts between 1981 and 1997. Moreover, they observed in their multivariate analysis that Mainline Protestants were significantly more likely to prevail in their claims that members of almost all other religious groups. Using a nine-category religious affiliation typology, they found that Catholics, Jews, Muslims, Baptists, sects and new religious movements as well as unspecified Christians were less likely than Mainline Protestants to prevail in their claims; the only claimants not significantly less likely to prevail were members of Native American faiths (Wybraniec and Finke 2001:437).

Articles by Sisk, Heise and Morriss (2004) and Claborn (2011) employed a different unit of analysis; rather than examining case outcomes, these studies considered judges’ votes. Examining cases in federal district courts and U.S. Courts of Appeals from 1986 and 1995, Sisk and his colleagues tested whether differences in judges’ characteristics affected how they voted on free exercise claims, though they found little evidence that this was the case. None of a long list of traits (including sex, race, political affiliation, education and employment background) made a difference in allowing them to predict how judges would vote, though they did observe differences based on religious characteristics of judges and the communities where they maintain their chambers (Sisk et al. 2004:491). Federal judges who were Jewish or belonged to Christian denominations outside of Mainline Protestantism, Catholicism or Baptist groups were significantly more likely to vote to uphold claims than members of Mainline Protestant groups. Likewise, the likelihood that judges would vote to uphold these claims increased with their community’s religious adherence rate.

The Sisk et al. (2004) study found significant associations in a multivariate framework between case outcomes and other factors. Using a nine-category typology to classify the substantive bases of cases, they found that cases concerning religious expression on public
property were significantly more likely than cases in their omnibus “other case type” category to be resolved in favor of free exercise claims. This finding offers confirmation of Way and Burt’s (1983) bivariate observation that courts tend to uphold public religious expression claims. Sisk and his colleagues also found that cases involving criminal misconduct were significantly less likely than these “other” cases to yield favorable outcomes for the claimants (2004:555). In addition, they also observed that claimants’ religions made a difference in free exercise cases. Catholics and Baptists were both significantly less likely to prevail in their claims than were members of most Christian denominations (a very broad category that includes Mainline Protestants, Pentecostals, Eastern Orthodox and Amish groups, among others), though no differences from this reference category are observed among Jews, Muslims, Native Americans and other non-Christian groups.

Finally, Claborn (2011) examined religious freedom cases in the state courts, the only study to focus explicitly on this level of analysis. However, this article analyzed outcomes only on a bivariate basis, and did not explicitly model influences on case outcomes (which Claborn operationalizes as judges’ votes). In addition, it does not appear that the universe of state court cases studied in this paper was limited only to free exercise claims; as he described in an earlier work analyzing the same dataset, Claborn’s analysis (2008:23-25) was “not confined to formal free exercise of religion decisions,” focusing more broadly on cases where claimants’ religious freedom is in question. This includes a variety of case types I have restricted from my analysis, including cases involving the striking of jurors at the trial level during voir dire due to their religious beliefs as well as cases involving the use of religious language in the public record by judges, prosecutors or defense attorneys.
Claborn studied case outcomes in the state courts from 1998 to 2005 and tried to distinguish between whether judges were more likely to vote in favor of upholding religious freedom claims after state attempts to “increase religious freedom,” which is conceived as the passage of state-level RFRA legislation and the employment of balancing tests in state courts. He categorized states according to whether they required these cases to be evaluated using some form of balancing test (which he scales based on the various levels of scrutiny different balancing tests require). Though he did not provide a statistical measure of association, Claborn (2011:624) noted that mandates to employ higher levels of scrutiny (i.e., more protective forms of balancing tests) were significantly associated in his analysis with higher rates of judge votes for free exercise claimants. This article was also the first to consider that case outcomes may be affected by the system states use for selecting judges, though he only reported distributions of outcomes based on judicial selection systems among states that had no defined standard for evaluating these claims (632).

In addition to these studies by sociologists and political scientists, a number of legal scholars have reviewed substantial numbers of free exercise decisions and reported outcomes. Arnold (1997) attempted to locate published opinions at all levels that included decisions on religious freedom claims made under Smith or the Religious Freedom Restoration Act; he located 174 such cases as of 1997. Remarkably, Arnold observed nearly identical rates of claimant success in cases decided under Smith as in cases decided under RFRA; in both cases, the success rate was approximately 45 percent (1997:152). This success rate is consistent with the rate among later cases reported by Claborn in 2011, but diverges greatly from success rates reported by Ryan (1992) and Lupu (1998), albeit over different periods of time. Reviewing the 97 claims in federal courts of appeals he located from 1980 to 1990, Ryan (1992:1416-1417)
observed that only 12 of the claims were upheld, a success rate of approximately 12 percent. Lupu (1998:590-591) reviewed the 168 published decisions at the state and federal levels made on the basis of the Religious Freedom Restoration Act prior to 1997. Only 25 of the 168 free exercise claims in these decisions were upheld, a success rate of about 15 percent. The dramatic differences in success rates reported here underscore a larger issue in this literature, which is that the success rates in all these studies vary a great deal, even within the same temporal window. Given that researchers’ case selection strategies are often rather opaque, it is important in the present analysis to be clear and transparent about how cases were selected for analysis; differences in case selection may account for some of the wide variation in free exercise claimants’ success rates in the courts seen in other studies.

Two relatively consistent findings emerge from the literature on outcomes of free exercise disputes in the courts. First, studies generally find that the year a claim was heard makes a difference in whether the claim was upheld or rejected. Cases heard during the period when the rational-basis test dominated free exercise jurisprudence (1990 to 1993) were generally less likely to yield successes for claimants than during periods when balancing tests were more prevalent. Studies often measured this effect by controlling for the legal time period in which the case was heard. Second, there appear to be differences in outcomes based on claimants’ religious affiliations; all studies that examined this except for Way and Burt (1983) found that mainline Protestants fared better than other groups.

**Dissertation Hypotheses**

Given the opportunity structure claimants face in the courts as described in this chapter, as well as the patterns of findings found in past literature, what are the factors that influence
whether free exercise claims are upheld or rejected? Furthermore, is it appropriate to assume that precedents or legislative mandates will be implemented and will lead to the use of the mandated form of legal test in these cases? Though a number of previous studies explored the former question, no study I am aware of has explored the latter. Using the hypotheses that follow, I attempt to answer these questions using the legal positivist and legal realist orientations toward the law explored in Chapter 1. In general, hypotheses grounded in the positivist view are based on the premise that variations in outcomes are the result of neutral application of the law. These hypotheses do not address the agency of decision-makers to act in ways not anticipated by strict interpretation of the law as written. On the other hand, the insights of legal realism suggest that accounting for judges’ attributes and aspects of the contexts in which cases are decided can help us understand how decisions are made (Cotterrell 1992:218). The realist hypotheses offered below are informed by the proposition that judges act based upon their attitudes as well as out of a desire to maximize their own self-interest.

I. Positivist Hypotheses

First, I propose a number of hypothesized relationships based on legal positivist logic. A number of the hypotheses below concern each of my two overarching research questions; hypotheses about variation in legal test use end with the suffix “-A,” while hypotheses about variations in case outcomes end with “-B.”

**Hypothesis One:** Claimants should be more likely to receive favorable decisions when courts adjudicate their claims using a balancing test instead of a rational-basis test.
Judicial tests that balance burdens on religious exercise against the weight of the government’s interest in imposing those burdens should pose a lower degree of difficulty for claimants than rational-basis tests. Justice Sandra Day O’Connor acknowledged this in her concurring opinion in the Smith decision, arguing that failing to employ the compelling interest test in Smith and other free exercise cases would place claimants at a greater disadvantage, especially members of minority faiths:

“Finally, the Court today suggests that the disfavoring of minority religions is an "unavoidable consequence" under our system of government, and that accommodation of such religions must be left to the political process…. In my view, however, the First Amendment was enacted precisely to protect the rights of those whose religious practices are not shared by the majority and may be viewed with hostility…. The compelling interest test reflects the First Amendment’s mandate of preserving religious liberty to the fullest extent possible in a pluralistic society. For the Court to deem this command a ‘luxury’… is to “denigrate ‘[t]he very purpose of a Bill of Rights’” (494 U.S. 872 [1990]).

As described earlier in the chapter, the compelling interest test as articulated in the court’s Sherbert v. Verner decision requires claimants to demonstrate that their ability to practice their religion has been burdened substantially; once that threshold is met, the government must demonstrate that it had a compelling interest in doing so and used the least-restrictive possible means of achieving its interest. Other state-level balancing tests also allow judges to consider
the severity of the burdens imposed on claimants’ religious practice. In contrast, rational-basis tests permit nothing of the sort; free exercise claims under this type of test fail if the government demonstrates that the offending law or policy is rationally related to a legitimate government interest. Given the differences in what the government (and by extension, the claimant) must prove under the two tests, we should expect that cases judged according some form of balancing test should generally yield more favorable results for claimants than cases judged according to a rational-basis test.

**Hypothesis Two:** Claims made by members of religious minority groups should be less likely to be upheld than claims made by members of more mainstream faiths.

As summarized above, previous research has generally found differences in free exercise case outcomes based on the minority status of claimants. This relationship has been tested in a number of different ways; Brent (1999, 2003) used a dichotomous mainstream/non-mainstream religious affiliation measure, while Wybraniec and Finke (2001) and Sisk et al. (2004) employed alternate multi-category affiliation typologies. Consistent with Black’s (1976:114) observation that outsiders will tend to be disadvantaged by the law, we should expect that members of more widely practiced and accepted faiths should fare better in pursuing their free exercise claims in court.

Adamczyk and her colleagues (2004:246) argue that the values and beliefs of majority religious groups are heavily embedded in American social norms. Because of this, such groups should be less likely to need to turn to the courts to resolve free exercise disputes, and when such disputes do arise, the behavior they wish to engage in should tend to deviate less from the corpus of socially legitimated behavior. As such, we should expect to see that members of more widely
practiced, less marginal faiths (such as most of the well-known Protestant denominations as well as Roman Catholicism) should be more likely to receive favorable decisions in their free exercise claims than more marginal groups with smaller populations in America (e.g., members of new religious movements and members of Eastern and Native American religions).

**Hypothesis Three-A:** After the passage of RLUIPA in 2000, claims involving prisoners’ religious rights or religious land use disputes should be more likely to be adjudicated using a balancing test than claims involving other substantive issues.

**Hypothesis Three-B:** Claims involving prisoners’ religious rights or religious land use disputes should be more likely to be upheld than claims involving other substantive issues after the year 2000.

The substantive natures of claims should influence the likelihood of favorable decisions for claimants based on much the same reason as undergirds Hypothesis One. The Religious Land Use and Institutionalized Persons Act (RLUIPA) singled out cases involving prisoners’ rights and land use (e.g., zoning) disputes for special treatment in federal and state courts; under RLUIPA, these cases should be adjudicated using the compelling interest test, a form of judicial balancing test. In fact, a number of state judiciaries employ the rational-basis test articulated in the Smith decision to adjudicate all free exercise claims except for these two types of claims. Consistent with the force of rule and precedent as stressed by legal positivist thought, judges will more likely than not employ the compelling interest test when evaluating these types of cases. As a result of RLUIPA, free exercise claims involving land use issues and those brought by prisoners should be more likely to be upheld than other types of cases during the period following the passage of RLUIPA (that is, after the year 2000).
Hypothesis Four: Judges will be more likely to use balancing tests to evaluate free exercise claims where state supreme courts and legislative actions mandate its use.

The last positivist hypothesis is perhaps the most fundamental to positivist reasoning: precedent should predict future action. Where the court of last resort has previously interpreted its state’s free exercise provisions to require the use of a balancing test in evaluating claims, judges will more likely than not follow this precedent and use this type of test themselves. This should also be true in states whose legislatures have passed “mini-RFRA” legislation codifying the use of a balancing test.

II. Realist Hypotheses

In addition to the hypothesized relationships described above, I also draw upon legal realist theory to hypothesize variations in legal test use and case outcomes based on more than neutral application of the law. Judges’ characteristics and their imperative to maximize their own self-interest should also influence both the likelihood of balancing test use in deciding free exercise cases and the outcomes of these cases themselves.

Hypothesis Five-A: Use of a balancing test should be less likely as the Republican composition of the panel of judges who hear claims increases.

Hypothesis Five-B: Claimants should be less likely to receive favorable decisions as the Republican composition of the panel of judges who hear their claims increases.

As described in Chapter 1, Republican-affiliated judges tend to be less likely to uphold civil liberties claims (Segal and Cover 1989; Marshall and Ignagni 1994). The latter of the two
hypotheses above assumes that this relationship will hold for the specific form of civil liberties claim examined in this dissertation: the free exercise of religion. Moreover, I hypothesize that given the ability to choose the form of legal test used in evaluating claims, judicial panels composed of greater proportions of self-identified Republicans or Republican appointees will be more likely to opt for a rational-basis test, the type of test less favorable to civil liberties claims. Though libertarianism and political conservatism are by no means incompatible, I expect to find that judges with more conservative ideologies will act in ways that are less protective of religious free exercise.

**Hypothesis Six-A:** In states where judges do not require public approval to retain their seats, claims made by members of minority faiths claims should be more likely to be adjudicated using a balancing test than in states where judges face popular referenda.

**Hypothesis Six-B:** In states where judges do not require public approval to retain their seats, claims raised by members of minority faiths should be more likely to be upheld than in states where judges face popular referenda.

One factor that has received little attention in studying the outcomes of free exercise cases is the way judges retain their seats on the bench. States use one of three broad classes of systems for retaining their appellate judges. In some states, judges may keep their seats through competitive elections that pit them against other candidates. In other states, judges face non-competitive retention elections, in which voters choose whether or not they should continue in their office for an additional term. Finally, a third class of states does not give citizens a direct voice in whether judges should retain their seats. Judges in these states may be appointed to the bench for life or else be retained by the vote of a legislative body or advisory commission. In
other words, there are some states in which the public has a direct say in whether or not judges can continue in their positions, and there are some states in which the public does not have a direct voice in this matter.

Consistent with Posner’s (1993) and Schauer’s (2000) assertion that judges are rational actors who want to maximize the benefits they receive, it makes sense to examine judicial behavior in the context of whether what judges do is the subject of public scrutiny through the electoral process. Even non-competitive retention elections should influence how judges behave; in their survey of judges who were subject to retention elections, Aspin and Hall (1994:312) found that a majority believed that voters were aware of judicial performance and relied on performance as the primary determinant influencing their votes. Given the assumption that judges generally want to retain their seats, I hypothesize that they will be less likely to act in ways that support minority faiths which may be less popular among the electorate, provided that they will face the electorate at a future point in time. This may be seen in their choice of judicial test for evaluating claims brought by religious minorities as well as in how they rule on these claims. On the other hand, judges who are not accountable to the public through popular referenda may be better insulated from these kinds of political considerations and may be more inclined to support claims brought by religious minorities, both via the test they use to evaluate these claims and in how they rule on the claims.

**Hypothesis Seven-A:** Free exercise claims will be more likely to be evaluated using a balancing test than a rational-basis test as the population of religious minorities in the immediate area where the case is heard increases.
Hypothesis Seven-B: Free exercise claims will be more likely to be resolved in favor of the claimant as the population of religious minorities in the immediate area where the case is heard increases.

Another way that judges’ attitudes might influence their decision-making is via their exposure to members of religious minorities in their communities. Social psychologists have long suspected that more sustained contact with members of out-groups can reduce levels of prejudice and build understanding (Allport [1954] 1979). Sisk and his colleagues (2004) investigated the question of whether the religious context of the community where a case is heard influences its outcome. They hypothesized that judges who come into contact with members of a wide array of religious faiths might be more likely to tolerate religious practices outside the mainstream (i.e., those that might not be protected by law). They found that greater likelihood of a favorable decision at the federal level was associated with higher religious adherence rates in the counties where cases were heard. Wybraniec and Finke (2001) also assessed the effect of regional context on the outcomes of free exercise cases, though they did not observe differences between outcomes of cases heard in the Midwest and those heard in either the religiously diffuse West or the heavily religious South. Nevertheless, broader exposure to members of religious minorities might influence judges to be more tolerant of their religious behavior in free exercise disputes, and it may also make them more inclined to resolve these disputes using a balancing test.

Summary

This chapter reviewed the development of two competing frameworks for understanding the scope of constitutional free exercise protections in America, discussed findings from
previous studies on the outcomes of free exercise cases, and introduced the hypotheses that I test in the chapters to come. The central question at the heart of the debate over free exercise in America is whether or not constitutional guarantees permit judges to give religious exemptions to otherwise neutral laws. The Supreme Court appeared to argue that this was the case from the 1960s through to 1990, but it took the opposite position in its *Employment Division v. Smith* decision. This reversal set off a flurry of activity that led to the patchwork opportunity structure for free exercise claimants in the state courts that exists today.

The dissertation models differences in how state appellate courts interpret free exercise protections explicitly. Previous studies indicate that the time period in which a claim was heard is a significant predictor of the claim’s success or failure, and this may be attributed largely to the different interpretations of free exercise rights that dominated in different periods. A comparison of state appellate courts, which employ both modes of interpretation, allows us to assess the effects of legal test use directly and in the present day. A number of additional hypotheses drawing upon legal positivist and realist logics were also reviewed in this chapter, and these hypotheses are tested later in the dissertation. In the chapters that follow, I discuss my data collection and analytic strategies, and I present and discuss the results of empirical tests of the hypotheses presented in this chapter.
CHAPTER 3: DATA AND METHODS

To test the hypotheses described in the previous chapter, I conducted a content analysis of all published state appellate court decisions from mid-1997 through the end of 2011. I also brought together data from a number of additional sources, including the Pew Forum’s U.S. Religious Landscape Survey, the U.S. Religion Census and CQ Press’ Judicial Staff Directory in order to perform my hypothesis tests. This chapter reviews the procedures used in conducting the content analysis as well as the specific measures and analytical techniques used in testing my hypotheses. I begin by discussing the procedures used to conduct the content analysis of court decisions. Next, I describe the measures used in testing the hypotheses as well as the data sources of measures not derived from the content analysis. Finally, I describe the quantitative analytic techniques I employed in testing my hypotheses.

Content Analysis of Court Decisions

A substantial amount of the data analyzed in this dissertation is drawn from content analysis of court decisions dealing with disputes over the free exercise of religion. Values for the two dependent variables (case outcomes and the legal tests judges used in each case) as well as the substantive basis for the free exercise claims were derived directly from this coding effort. In addition, values for all other independent variables relied on information from the court decisions. For instance, the proportion of the judicial panel comprised of Republican judges in each case was measured by recording the names of all judges listed in the opinion and obtaining their political affiliations from other sources. Likewise, I derived my measure of claimants’ religious affiliation by recording this affiliation from the court decisions, classifying that affiliation into one of several groups, and identifying the size of that group’s population as a
proportion of the total state population. This section reviews the procedures I used for locating and coding court decisions.

I. Using LoislawConnect to Locate Decisions

I located and coded court decisions from January to March of 2012 in order to conduct my analysis. These decisions represent the universe of all published free exercise cases ruled upon in state appellate courts (i.e., intermediate courts and courts of last resort) between June 25, 1997 and December 31, 2011 which were not remanded to a lower court for further consideration. The window of analysis begins on the same date as the Supreme Court’s pivotal ruling in City of Boerne v. Flores. As reviewed in Chapter 2, the Court’s ruling in this case invalidated federal legislation that had mandated the use of the compelling interest test (a form of balancing test) at all levels when judges heard free exercise claims based on the First Amendment. Flores established the paradigm that First Amendment free exercise claims heard at the state level should be evaluated using the rational-basis test articulated in the Court’s 1990 decision in Employment Division v. Smith. In order to understand the present-day factors influencing case outcomes and judicial test selection, I only consider the decisions made after Flores in mid-1997.

In order to ensure that I included only cases at the appellate level in my analysis, I relied on court structure charts produced by the Court Statistics Project of the National Center for State Courts, a non-profit organization funded in part by the U.S. Department of Justice’s Bureau of Justice Statistics.18 Trial court opinions were excluded from analysis due to their lack of availability; only a small number of states (e.g., Connecticut) publish opinions from courts at this

---

18 Court structure charts for each state are available at the website of the Court Statistics Project (http://www.courtstatistics.org/Other-Pages/State_Court_Structure_Charts.aspx).
level. As a result, only decisions from those courts listed as intermediate appellate courts or courts of last resort were included in the analysis.\textsuperscript{19} The universe of analyzed cases only includes those in which the court made a ruling on free exercise grounds; this excludes cases involving multiple types of claims (including a free exercise claim) in which the free exercise claim was not explicitly ruled upon, as well as cases that were remanded to a lower court.

I located the universe of court decisions meeting these criteria using LoislawConnect, an electronic research database service owned by the legal publisher Wolters Kluwer. LoislawConnect is one of several electronic research databases used by scholars and legal professionals and scholars to search for case law; the most prominent and best-known of these resources are Westlaw and LexisNexis. Among LoislawConnect’s features is a case law database for every state and the District of Columbia; the database contains the text of opinions from at least 1950 for most states (Sanborn 2000). Use of this database for legal research is becoming widespread in large law firms as a lower-cost, primary source alternative to Westlaw and LexisNexis.\textsuperscript{20} I accessed LoislawConnect through the Pennsylvania State University library system’s online database holdings.

I verified that using this database would not suppress the number of cases in my analysis by searching LoislawConnect and LexisNexis, a better-known alternative database, for free exercise cases. Searches for state cases in these databases containing the terms “religio*” and “exercis*” (allowing both root words to have different suffixes, e.g., “religion” and “religious”)

\textsuperscript{19} In the example of Pennsylvania, decisions made by the state Supreme Court, Commonwealth Court and Superior Court were coded, while decisions made by trial courts (Courts of Common Pleas, Philadelphia Municipal Court, Philadelphia Traffic Court and Magisterial District Courts in Pennsylvania) were not. See \url{http://www.courtstatistics.org/Other-Pages/State_Court_Structure_Charts/Pennsylvania.aspx} for the Court Statistics Project’s court structure chart for the state of Pennsylvania. If a state had no court at the intermediate appellate level (e.g., Maine and Wyoming), I only coded decisions from the state’s court of last resort.

\textsuperscript{20} In a survey of 162 law firms, 84\% of which employed more than 75 attorneys, Justiss (2011) found that LoislawConnect was the most commonly used alternative to the two dominant services; while 38\% of firms used either Westlaw or LexisNexis for conducting legal research, 16\% used LoislawConnect in its former incarnation as Loislaw.
yielded a greater number of cases in LoislawConnect than in LexisNexis for all but four states.\textsuperscript{21} I concluded based on the breadth of LoislawConnect’s results that use of this database would not systematically exclude free exercise cases from analysis.

\section*{II. Case Selection}

The final search term employed in locating decisions was developed using a two-stage process. To generate a search term that would yield a maximal number of true free exercise decisions while minimizing the number of false positives, I first queried LoislawConnect’s database of appellate court decisions in five states (Alabama, Alaska, Arizona, Arkansas, and California) using a very broad search term.\textsuperscript{22} This yielded a total of 913 cases, only 72 of which concerned disputes over the free exercise of religion. In the second stage of this process, I replaced my original search term with a series of “or” terms (i.e., a system of terms that would yield cases containing any of the specified terms). The final search term reproduced the bona fide decisions from the earlier search while returning 45 percent fewer total opinions.\textsuperscript{23} This search term produced 3,828 state-level cases across the fifty state court systems during the period of analysis. The large majority of these opinions were false positives, but 453 of these opinions contained disputes over religion.

I conducted a content analysis on these 453 decisions using the coding instrument provided in Appendix A. After coding them, I limited the number of decisions in my analysis to

\textsuperscript{21} The only states in which LexisNexis outperformed LoislawConnect were California, Maine, Michigan and Montana.

\textsuperscript{22} The LoislawConnect search term I began with was “religio* & exercis*”, which yielded all cases containing words that included both the consecutive letters “religio” and “exercis.” This was done to account for variations in the forms of the key root words (e.g., “religion,” “religious” and “religiosity” for the first root and “exercise” and “exercising” for the second).

\textsuperscript{23} The final LoislawConnect search term used to locate cases is as follows: “free exercise” | “religious exercise” | “exercise of religion” | “religious freedom” | “rluiwa” | “religious land use” | “rfra” | “freedom restoration” | “religious right*” | “freedom of religion” | “religion clause*” | “ministerial exemption” | “ministerial exception” | “ecclesiastical rule” | “(exempt* & religio*)” | “(privilege* & clergy*)” | “(exercis* near5 religion).”
a total of 225 decisions. For a decision to be included in the final 225 free exercise cases, it had to meet the following criteria:

a) The claimant alleged that their right to the free exercise of their religion was or would be burdened by government policy, action or law;

b) Either the appellant or respondent in the case alleged that their free exercise rights were or would be burdened, but not both parties (in other words, competing free exercise claims made by opposing parties were not included); and

c) The burden did not pertain to a benefit that religious groups already enjoy under the law but which secular groups do not (e.g., religious tax-exempt status or exemption from anti-discrimination laws on religious grounds).

The majority of the decisions that were coded but not included in the final analysis failed to meet the final criterion. Many of these decisions pertained to whether religious groups were entitled to tax exemptions for certain pieces of property, an issue adjudicated in the courts using a different set of standards (see Walz v. Tax Commission of the City of New York, 397 U.S. 664 [1970]; Hernandez v. Commissioner, 490 U.S. 680 [1989]; Jimmy Swaggart Ministries v. Board of Equalization of California, 493 U.S. 378 [1990]). Other decisions were not included because they failed to meet the second criterion of “no competing claims.” These cases tended to pit rival factions of a single religious organization against one another over the organization’s operation.

A brief word is warranted here on the decision to include only published decisions in the final analysis. A substantial number of federal and state court opinions are not released for publication, and the reasons for doing so vary by jurisdiction. Generally speaking, unpublished

---

24 As shown in Table 4 in Chapter 4, I did include two cases involving tax disputes, but both of these cases explicitly contained claims made under a constitutional free exercise clause, and unlike most religious tax exemption cases, the claims in both cases were evaluated using one of the two types of legal tests I discuss throughout the dissertation.
opinions are those that judges did not regard to have significant precedential value. Publication rates vary widely by jurisdiction, and at the federal appellate level, no circuit releases more than half of its opinions for publication (Bayer 2009). Nevertheless, some unpublished decisions are available through LoislawConnect, though their availability varies widely across states. Researchers who analyze appellate court decisions must decide whether or not to include any unpublished decisions in their analysis.

Both options run the risk of introducing potential biases into the analysis. The decision to exclude unpublished decisions may overstate claimant successes in free exercise cases, for as Sisk and his colleagues (2004:535) argued, published opinions may be less likely to address frivolous, easily dismissible claims. On the other hand, including unpublished decisions is likely to introduce biases, since we cannot know if the small number of unpublished decisions available through primary source databases like LoislawConnect is representative of all unpublished decisions. Decisions of this kind are not likely to have been made available at random; more recent unpublished decisions and decisions from certain states (e.g., Minnesota and Ohio) are more likely to be available than others.

I follow the approach of Sisk and his colleagues (2004:535) in their study of religious freedom cases at the federal level and include only published decisions in my analysis. Research suggests that beyond the matter of claims’ viability, few external factors predict whether cases will be published; this suggests that my analysis would not be substantially biased by excluding available unpublished decisions. Merritt and Brudney (2001), for instance, found little difference in publication likelihood based on judges’ political affiliation, the gender composition of the panel of judges and other factors. Moreover, factors they found to predict publication, such as the age of judges and whether they graduated from elite law schools, would presumably
exert little systematic influence on their decision-making on free exercise claims in the state courts. Given these findings and prior research strategies in this area, I exclude unpublished decisions from my analysis.25

The coding instrument given in Appendix A provides forty-nine different measures that I coded for each decision, including fourteen dichotomous variables that can be combined to form a single measure of the free exercise dispute’s substantive basis. A number of the measures included in the codebook are not included in the statistical models reviewed in Chapters 5 and 6, namely the measures identifying whether cases cited major past opinions or legislation. These measures were useful in confirming values for other variables but are not included in the present analysis.

III. Measurement Reliability

To assess the reliability of the coding instrument, a second coder reviewed and coded all selected court decisions from every fourth U.S. state in alphabetical order. The coder reviewed cases from Alaska to Wyoming and again from Arkansas to Mississippi, for a total of 19 states or 38 percent of all U.S. states.26 I compared the dataset I generated by my coding against the

---

25 It should also be noted that because I treated individual decisions as the unit of analysis, I decided to include decisions on individual cases when they were heard both by an intermediate court and a court of last resort. For example, I included the Texas Court of Appeals’ 2005 decision in Barr v. City of Sinton (295 S.W.3d 334) as well as the Texas Supreme Court’s 2009 decision in the same matter (295 S.W.3d 287). However, I believe that any biases that might be introduced using this method are minimal, as only 13 such case dyads were included in my final analysis. In all, I analyze 160 decisions from intermediate appellate courts and 65 decisions from state courts of last resort, 15 of which were heard in states with no intermediate appellate court.

26 State court opinions were read and coded by the second coder for the states of Alaska, Arkansas, Colorado, Delaware, Georgia, Idaho, Indiana, Kansas, Louisiana, Maryland, Michigan, Mississippi, Montana, New Jersey, North Dakota, Pennsylvania, Tennessee, Virginia and Wyoming. Alaska was chosen as the first state to code after the coder was trained by coding decisions from Alabama, the first state that appears in an alphabetical list of U.S. states.
Table 1: Intercoder Reliability of Selected Measures

<table>
<thead>
<tr>
<th></th>
<th>% Agreement</th>
<th>Cohen’s Kappa</th>
<th>Krippendorff’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of Legal Test</td>
<td>82.54%</td>
<td>.725</td>
<td>.726</td>
</tr>
<tr>
<td>Favorable Decision</td>
<td>83.33%</td>
<td>.623</td>
<td>.624</td>
</tr>
<tr>
<td>Case Type</td>
<td>86.51%</td>
<td>.844</td>
<td>.845</td>
</tr>
<tr>
<td>Level of Court</td>
<td>97.62%</td>
<td>.948</td>
<td>.949</td>
</tr>
<tr>
<td>Citation of Smith</td>
<td>95.24%</td>
<td>.892</td>
<td>.892</td>
</tr>
<tr>
<td>Citation of Sherbert v. Yoder</td>
<td>97.62%</td>
<td>.913</td>
<td>.913</td>
</tr>
<tr>
<td>Citation of Religious Freedom Restoration Act</td>
<td>96.83%</td>
<td>.890</td>
<td>.890</td>
</tr>
</tbody>
</table>

The coding performed by the author and reliability coder suggests an acceptable level of intercoder reliability. The percent agreement between the two coders on the selected measures is relatively high, registering above the 80 percent threshold for each of the variables provided above. The most crucial variables in this table are the first three listed (whether a decision was favorable to the claimant, the type of legal test used to evaluate the claim, and the substantive nature or “type” of the case), as these are the most important of the measures in Table 1 for second coder’s dataset to assess intercoder reliability. Using Freelon’s (2010) ReCal2 utility, I calculated values of three intercoder reliability statistics: percentage agreement, Cohen’s kappa (κ), and Krippendorff’s alpha (α). Table 1 above provides values for each of these three measures of intercoder reliability on selected measures.

ReCal2 is available for free public use at [http://www.dfreelon.org/utils/recalfront/](http://www.dfreelon.org/utils/recalfront/).
testing the hypotheses reviewed in Chapter 2. Coders were in agreement 83 percent of the time on whether claimants received favorable decisions and on the form of legal test used, while agreement on the substantive basis of each claim was higher (approximately 87 percent agreement).

In addition to the percentage agreement measure, I calculated two methods of intercoder reliability using ReCal2 that take into account the probability of agreement occurring by random chance: Cohen’s kappa (Cohen 1960) and Krippendorff’s alpha (Hayes and Krippendorff 2007), which both have ranges from 0 to 1. As with the measure of percentage agreement, interpretation of these measures regarding whether scores represent high or low levels of reliability depends on the use of guidelines rather than objective standards. Landis and Koch (1977:165) argue that kappa values between .61 and .80 represent substantial agreement and kappa values between .81 and 1.00 represent almost perfect agreement. By this standard, the degree of intercoder reliability on the measures of successes (κ = .623), legal test (κ = .725) and case type (κ = .844) can be considered as substantial. Krippendorff’s alpha differs from Cohen’s kappa in that it accounts for systematic disagreements between coders more consistently (Hayes and Krippendorff 2007); however, with this matrix of data, the difference between values on this measure and Cohen’s kappa is negligible. In short, though the calculated values of intercoder reliability statistics do not suggest perfect agreement between coders, they do suggest that the way in which the data were coded would be relatively consistent regardless of coding decisions made by different raters.

---

As will be described below, claimants’ religious affiliations were initially coded using an open-ended measure because of considerable differences in the specificity of affiliations provided across court decisions; affiliations were then coded based on additional web-based investigation into the denominational backgrounds of the claimants where applicable as well as reference to Melton’s Encyclopedia of American Religions (2009). Because of this strategy, the level of intercoder reliability for the measure of claimants’ religious affiliations was not assessed.
In the following section, I describe the measures I used to test my hypotheses regarding my two research questions. Where measures were not based on my content analysis, I provide short descriptions of each of their data sources. I discuss my two dependent variables and then describe the array of independent and control variables I employ in my analysis.

MEASURES

I. **Dependent Variables**

The dependent variable in each of the models in Chapter 5, and the final outcome variable as organized in the structural equation models in Chapter 6, measures whether or not claimants received a favorable decision from a state appellate court. In other words, it is a dichotomous measure indicating whether or not any part of the free exercise claim was upheld. Any ruling not completely in favor of the opposing party (usually a representative of the state) was coded as a favorable decision. An example of this kind of decision is *Hyde v. Fisher* (203 P.3d 712 [Idaho 2009]), an Idaho case in which a practitioner of a traditional Native American faith requested access to religious materials and permission to practice sweat lodge and smudging ceremonies while incarcerated. Applying the compelling interest test to Hyde’s claims under both the Religious Land Use and Institutionalized Persons Act (RLUIPA) and Idaho’s state RFRA legislation (FERPA), the court in this case ruled as follows:

“The district court correctly ruled that [the Idaho Maximum Security Institution’s (IMSI)] complete ban on sweat lodge ceremonies and implementation of a personal property policy does not violate Hyde’s constitutional rights conferred upon him by RLUIPA and
FERPA. However, we conclude that [the Idaho Department of Correction] has not demonstrated that completely banning smudging ceremonies at IMSI is the least restrictive means of furthering the compelling government interest of safety and security at the institution.” (emphasis in original)

Having found that denying the claimant access to religious rites and property imposed a substantial burden on their free exercise rights, the court considered whether this burden was justified by a compelling government interest, and whether imposing the burden was the least restrictive means of furthering that interest. The safety and security of the correctional institution where the claimant was incarcerated were found to be compelling interests. However, the court held that the prison could pursue these interests without needing to ban one of the rituals the claimant requested: the smudging ceremony, which involves bodily purification through inhaling and coming in contact with smoke from burned elements. As a result, this case was coded as a favorable decision for the claimant, even though they did not receive everything they had asked for in their claim (e.g., the ability to hold a sweat lodge ceremony in prison).

The dependent variable tested in the first half of Chapter 6, which is considered an independent variable in Chapter 5, measures the type of legal test judges used to adjudicate each claim. I initially coded legal tests using one of three categories: balancing tests, rational-basis tests and other types of tests. Decisions based on the latter category were removed from the final analysis as described earlier in the chapter, which reduced the number of observations in the analysis from 453 to 225. Most of the cases adjudicated using the third type of test did not meet one of my three inclusion criteria; they frequently concerned disputes over tax exemptions or other special benefits, or else involved disputes between multiple free exercise claimants.
I identified rational-basis tests using the coding instrument according to whether judges interpreted free exercise protections as follows: as long as a law is rationally related to a government interest and/or does not discriminate on religious grounds, no free exercise exemptions are possible. To aid in identifying the type of legal test judges used, the coding instrument lists key precedents that are notable examples of each type of test. *Employment Division v. Smith* is the most prominent and frequently-cited example of a precedent cited when judges use a rational-basis test, but there are other notable examples:

- *Turner v. Safley*, 482 U.S. 78 (1987) and *O’Lone v. Estate of Shabazz*, 482 U.S. 342 (1987). These two 1987 Supreme Court decisions set the precedent for evaluating disputes over the religious rights of the incarcerated. *Turner v. Safley* established that prison officials needed only to justify their restrictions on inmates’ First Amendment rights on the grounds of a rational, legitimate penological interest. Review under a more protective standard, such as the compelling interest test, was not required even prior to the Court’s decision in *Employment Division v. Smith*. *O’Lone v. Estate of Shabazz* upheld the rational-basis test established in *Turner v. Safley* and made it applicable to disputes over prisoners’ religious free exercise rights. The rational-basis test employed in both cases functioned as precedent in the state courts until the passage of the Religious Land Use and Institutionalized Persons Act (RLUIPA) in 2000; some states continued to rely on these precedents even after RLUIPA was enacted.  

- *Church of the Lukumi Babalu Aye v. City of Hialeah*, 508 U.S. 520 (1993). This Supreme Court decision upheld the rational-basis test used in *Smith* while also demonstrating how claimants might prevail under such a test. This case concerned a city

---

ordinance that banned ritual animal sacrifice in the city limits of Hialeah, Florida; the ordinance would have prevented a Santeria church which engages in this practice from operating legally in the city limits. The Court applied its Smith test in evaluating this dispute and ruled that the city ordinance was not neutral toward religion and indeed targeted religious groups for sanction. Since the government failed to meet the threshold for denying cases under the rational-basis test, the Court then applied the compelling interest test to the facts of the case and concluded that the ordinance was unconstitutional. A number of state courts ruling on free exercise disputes have relied specifically on the City of Hialeah precedent when employing the rational-basis test upheld in this case.\textsuperscript{30}

The coding instrument identifies balancing tests as those requiring judges to consider burdens allegedly imposed on claimants’ religious rights by law or policy. Two Supreme Court precedents that judges in state courts frequently cite when justifying the use of a balancing test, as described in Chapter 2, are Sherbert v. Verner (1963) and Wisconsin v. Yoder (1972); these cases are listed in the coding instrument as examples of this type of test.\textsuperscript{31} Likewise, the coding instrument indicates that cases in which judges rely on the Religious Freedom Restoration Act (RFRA), RLUIPA and state-level “mini-RFRAs” to make their rulings are examples of balancing test use. In addition, references to the use of “strict scrutiny” or “compelling interest tests” in justifying a ruling indicate the use of a balancing test.

\textsuperscript{30} See Reid v. Arkansas Department of Human Services, 2011 Arkansas 187; Malicki v. Doe 814 So.2d 347 (Florida 2002); Spell v. Muhammad, 756 So. 2d 748 (Mississippi 2000).

\textsuperscript{31} For examples of cases that rely on one of these precedents to justify the use of a balancing test, see Conley v. Roman Catholic Archbishop of San Francisco, 102 Cal.Rptr.2d 679 (California 2000); McCready v. Hoffius, 586 N.W.2d 723 (Michigan 1999); and Alexander v. State, 282 S.W.3d 143 (Texas 2009).
II. Independent and Control Variables

a. Religious Affiliation and Minority Status

A substantial majority of decisions contained information about the claimants’ religious background. Some 187 of the 225 decisions included some nominal information on the claimant’s religious origin. This information varied in its specificity; some decisions provided the claimant’s specific religious denominational or congregational membership, while others recognized the claimant as a “Christian” or “Muslim.” Recognizing the disparities in the religious identification information provided in these opinions, I opted to have coders record claimants’ religious affiliations as provided in each opinion without using a forced-choice response item in the coding instrument. After the initial coding, I obtained further information about the nature of the claimants’ denomination or congregation by referring to Melton’s *Encyclopedia of American Religions* (2009) (especially regarding new religious movements) and/or by locating congregations’ or denominations’ websites or internet presence.

In order to test my hypothesis on the link between successful case outcomes and claims made by members of minority faiths, I classified each claimant for whom religious affiliation information was provided into one of fourteen affiliation categories. The classification scheme I used is based predominantly on the Pew Forum on Religion & Public Life’s (2008) typology of religious families (the Pew Forum variable FAMILY) in its U.S. Religious Landscape Survey. I preserved the following categories from this typology in my analysis: Catholics, Historically Black Protestants, Orthodox [Christians], Jews, Muslims, Buddhists, Hindus, and members of Other World Religions (e.g., Sikhs, Zoroastrians, Taoists and members of other Eastern faiths).
In addition, I combined separate categories for New Age groups, which include new religious movements, and Native American religions into a single category.

Categorizing Protestants who are not members of the historically Black Protestant tradition proved more challenging. I opted against relying on the two-category distinction made by both the U.S. Religious Landscape Survey and the RELTRAD classification scheme (Steensland, Park, Regnerus, Robinson, Wilcox and Woodberry 2000), which is often used in sociological studies of religion. The Evangelical/Mainline dichotomy used in both systems may elide over key signifiers of religious marginality among its member denominations. For instance, Evangelical Protestants, according to the Pew Forum’s classification, count Southern Baptists as well as a host of smaller, stricter groups among their membership. Given their broad representation in the U.S. population, Southern Baptists can hardly be considered a minority religious group; it does not make sense to classify all Evangelical Protestants as religious minorities in America.

Instead, I draw upon strict church theory to posit a religious marginality ranking across denominational Protestants. Iannaccone (1994:1191) surveyed religious scholars and asked them to rank Christian religious denominations by how strongly they emphasized a separate, stricter lifestyle and morality than is accepted in broader American society. His survey attempted to replicate the religious strictness scale first developed by Hoge and Roozen (1979) through the same strategy. The results of Iannaccone’s survey mirrored those of Hoge and Roozen quite strongly, with a Cronbach’s alpha value greater than .98. Based on the results of his survey, Iannaccone posited a four-category denominational classification according to each denomination’s strictness level: liberal, moderate, conservative and sect. His distribution of denominations by strictness level is provided in Table 2 on the following page.
Table 2: Denominational Protestant Categories

<table>
<thead>
<tr>
<th>Liberal Denominations</th>
<th>Moderate Denominations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episcopalian</td>
<td>Lutheran (ELCA)</td>
</tr>
<tr>
<td>Methodist</td>
<td>Reformed Church</td>
</tr>
<tr>
<td>Disciples of Christ</td>
<td>Presbyterian</td>
</tr>
<tr>
<td>United Church of Christ</td>
<td>American Baptist</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservative Denominations</td>
<td>Sects</td>
</tr>
<tr>
<td>Southern Baptist</td>
<td>Assemblies of God</td>
</tr>
<tr>
<td>Lutheran (Missouri Synod)</td>
<td>Church of Christ</td>
</tr>
<tr>
<td></td>
<td>Church of God</td>
</tr>
<tr>
<td></td>
<td>Jehovah’s Witness</td>
</tr>
<tr>
<td></td>
<td>Church of the Nazarene</td>
</tr>
<tr>
<td></td>
<td>Seventh-Day Adventist</td>
</tr>
<tr>
<td></td>
<td>Church of Jesus Christ of Latter-day Saints</td>
</tr>
<tr>
<td></td>
<td>Other Pentecostal groups</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>


I ranked denominational Protestant denominations from most to least strict using Iannaccone’s (1994) classification scheme. Next, I located each of the denominations provided in Table 2 in the Pew Forum’s DENOM classification system used in the U.S. Religious Landscape Survey, which contains a far larger number of denominations than Iannaccone’s scheme. The DENOM system groups denominations by tradition (e.g., the Southern Baptist Convention is coded as “Baptist in the Evangelical Tradition”). All other denominations under the same heading as a denomination in Table 2 were included in one of the four appropriate strictness categories; for example, all other denominations classified as “Baptist in the Evangelical Tradition” in DENOM were coded as “Conservative Denominations” using the strictness scheme in Table 2. Non-denominational Christian groups were treated somewhat differently. In a number of decisions, claimants belonged to congregations that identified

---

The full codebook, which categorizes a panoply of denominations by tradition within the DENOM variable, may be downloaded at [http://www.pewforum.org/FormPage.aspx?ekfrm=155](http://www.pewforum.org/FormPage.aspx?ekfrm=155).
themselves as explicitly non-denominational. Absent objective standards for classifying these non-denominational groups within any of the Pew Forum’s denomination-traditions (or within one of the four strictness categories), I elected to code them as “non-denominational Christians.”

Religious minority status can be conceptualized in multiple ways. Previous quantitative studies of free exercise disputes – and indeed, many studies that include religious groups as predictors – treat religious affiliation as a nominal variable. These studies usually compare effects of membership in different dummy-coded religious groups against a single group membership reference category. This approach poses problems for studies of religious marginality, in that it may force researchers to assign minority as well as majority status to various religious groups when these statuses may be controversial. For instance, Catholicism was marginal enough only five decades ago that John F. Kennedy had to reassure the nation he would not take commands from the pope if he were elected the country’s first Roman Catholic president. Catholicism now constitutes the single largest religious denomination in America, and two-thirds of the justices on the Supreme Court in 2013 were Catholic. On the other hand, Episcopalians and Congregationalists are among the groups with the oldest historical legacies in America (Finke and Stark 2005:56), so much so that they are classified as “Mainline” Protestants in Steensland et al.’s (2000) RELTRAD scheme. However, less than one-fifth of all Americans now belong to a Mainline Protestant congregation in the 21st century (Pew Forum 2008).

While supplemental models in Chapter 5 compare the likelihoods of claimant successes among different religious groups using the more traditional nominal approach, I prefer to

In sum, each claimant was categorized as belonging to one of the following fourteen groups: Liberal Protestant, Moderate Protestant, Conservative Protestant, Protestant Sect, Black Protestant, Roman Catholic, Orthodox Christian, Jewish, Muslim, Buddhist, Hindu, New Age/New Religious Movement/Native American Religion, Other World Religions, and non-denominational Christian. Liberal and Moderate Protestant populations were combined in the analysis in later chapters.
examine the effect of religious marginality on the likelihood of success in a different manner. This study treats religious minority status as a continuous variable by using religious demographic statistics available from the Pew Forum’s U.S. Religious Landscape Survey (2008). Once I categorized a claimant into one of fourteen religious affiliation categories, I obtained the proportion of the total state population who belong to the same affiliation category as the claimant from U.S. Religious Landscape Survey data.\textsuperscript{34} Next, I transformed the resulting distribution of proportions using a logit transformation. The untransformed distribution has a markedly positive skew due to the overrepresentation of new religious movements and other very small faiths among the observations, and this is normalized by performing this transformation. In this way, religious minority statuses can be compared along a single continuous dimension.

b. Substantive Nature of Dispute/Case Type

The factual nature of the claims was also coded from the text of the decisions. I classified cases using a typology that I generated by referring to earlier free exercise studies (Way and Burt 1983:659; Sisk et al. 2004: 559-60) and by perusing free exercise decisions before coding. As shown in Appendix A, I initially classified decisions into one of fourteen discrete substantive categories. To assess whether the measure was coded reliably, I collapsed the fourteen dichotomous category variables into a single, fourteen-category nominal variable and calculated reliability statistics for this measure across coders (as described earlier in this chapter); the results of this analysis demonstrated a substantial degree of intercoder reliability on this measure ($\kappa = .844$). The case type variables were collapsed into four basic categories for the

\textsuperscript{34} In the final analysis, proportions of Liberal and Moderate Protestants in states’ populations were combined, as there is little difference between the cultural statuses of these two groups in contemporary society. However, proportions from the remaining twelve groups were calculated independently of one another.
analysis in subsequent chapters: prison and land use disputes, disputes over education and child welfare, disputes over “positive” government policies and other types of disputes.

Given their unique status under the Religious Land Use and Institutionalized Persons Act (RLUIPA), I assigned prisoners’ rights and land use disputes to a single case type category. After the year 2000, claims of these two types were to be evaluated using the compelling interest test under RLUIPA. Disputes involving educational institutions and the welfare of children were likewise categorized together. The third category—disputes over positive government policies—contains cases involving claimants’ desire to not take certain actions that the government compels everyone to take. For example, many cases in this category concern claimants’ religious objections to obtaining government-issued identification. The key distinction between cases in this category and others is that these cases concern objections to mandatory positive action that governments require all citizens to take regardless of their religion. The final case type category includes other types of disputes, including disputes over the use of banned substances, unemployment compensation and religious exemptions from tort liability.\(^{35}\)

c. Judicial Characteristics

Two sets of hypotheses drawn from legal realist logics concern the likelihood of receiving favorable decisions in free exercise claims as a function of judges’ political affiliations and their desire to maximize their own self-interest. I needed to obtain judges’ political affiliations and information on how they retain their judgeships in order to test these hypotheses.

\(^{35}\) The frequency distribution of coded case types is provided in Table 4 in Chapter 4; the fourteen original case type categories are provided within the later four-category classification scheme. This table also lists two additional categories (Other Dispute: Exemption from Tort Liability and Other Dispute: Claim That Law is Unconstitutional) which were coded retroactively by examining cases that received a score of 1 on TYPEOTH (the case type variable for decisions that were not classifiable in other categories). However, the addition of these two categories to Table 4 does not alter the empirical analysis in subsequent chapters at all.
This information was not available through content analysis of court decisions and required me to seek out additional sources of data.

To obtain the proportion of Republican judges on each judicial panel, I first recorded the names of all judges who heard claims, which were generally provided in the text of decisions. Once the names of all judges were recorded, I obtained their political affiliations using a number of sources. The vast majority of judges’ affiliations were obtained from the year 1998 through 2012 editions of the *Judicial Staff Directory*, CQ Press’ annual publication that provides background information on judges at all levels. Where judges’ political affiliations were not available through the *Judicial Staff Directory*, I consulted *The American Bench*, an annual publication by Forster-Long that also provides this type of background information, as well as online editions of newspapers and political websites (namely judges’ former campaign sites and the websites of state political parties).36

One complication in obtaining these data is that governors in many states appoint judges to the bench who do not disclose their political affiliations. In these instances, I assigned the judges the political affiliation of the governors who appointed them. However, in a small number of states where judges are retained through nonpartisan elections and where they were not initially appointed by a political actor or body (e.g., Washington state), I was not able to record political affiliations unless they were disclosed by the judges. I measure the political

36 In addition to this measure of political affiliation, I had also intended to include a measure of the religious affiliations of the judges who had heard each claim. Though previous studies of judicial behavior suggest that religious affiliation can be a key factor in determining how judges will rule on certain types of cases (Songer and Tabrizi 1999; Sisk et al. 2004), I was not able to include this measure due to the dearth of available data on state appellate judges’ religious affiliations. A preliminary attempt at coding judges’ religious affiliations based on *The American Bench* and the *Judicial Staff Directory* yielded affiliations for only 15 percent of judges (nine out of 59 judges researched). This constituted an unacceptable level of missingness, and so I opted to drop this variable from my analysis.
composition of each judicial panel as the proportion of judges who self-identify as Republican or, when they do not identify a political affiliation, were appointed by a Republican governor.\footnote{A disadvantage of this approach is that it effectively categorizes the appointments of three governors who were not affiliated with a political party while in office (Jesse Ventura in Minnesota, Angus King in Maine and Lowell Weicker in Connecticut) as de facto Democratic appointees. However, only eight judges who heard any of these cases were (a) appointed by one of these governors and (b) had failed to disclose their political affiliations in either the \textit{Judicial Staff Directory} or \textit{The American Bench}. Together in consideration of the civil libertarian orientations of each of these three governors, this makes it unlikely that my approach would bias the results of analysis using this political affiliation measure.}

To explore the effect different systems of judicial retention may have on the likelihood of favorable rulings, I referred to the website of the American Judicature Society.\footnote{This map is available through the American Judicature Society’s website at http://www.judicialselection.us/} This site features an interactive United States map that describes how appellate judges in each state retain their seats. As reviewed in Chapter 2, states use three broad classes of judicial retention procedures: (1) competitive elections between judicial candidates, (2) non-competitive retention elections, in which voters decide whether or not judges should retain their seats, and (3) procedures that do not give the public a say in whether or not judges should be retained. The latter include permanent judicial appointments and retentions based on the approval of a judicial commission or state legislature. I coded each case according to the judicial retention system employed where the case was heard as described by the American Judicature Society’s website. Since my hypotheses about the effects of the retention systems on outcomes and use of legal test are based on claimants’ minority status, I test them by interacting judicial retention system indicators with the continuous measure of religious group population size described earlier in this section.

d. Other Variables

I include four additional measures in my models of case outcomes and use of legal test in subsequent chapters. First, I test my hypotheses about the effects of judicial exposure to...
religious minorities (regardless of the claimants’ affiliation) through a broad measure of local religious minority population. I rely on data from the 2010 U.S. Religion Census Metro Area data file (also known as the Religious Congregations and Membership Study, or RCMS; see Grammich, Hadaway, Houseal, Jones, Krindatch, Stanley and Taylor 2012). This dataset contains religious population data for every metropolitan statistical area in the country, while data on micropolitan statistical areas is available from the same source through the Association of Religion Data Archives’ U.S. Congregational Membership report pages.

I coded the local religious minority population where each case was heard in three steps. First, using information provided through content analysis of court decisions, I located the metropolitan statistical area in which the case was heard; I found each court location by searching each court’s official website. Next, I created a raw minority population size variable by adding the population of Orthodox Christians to the population of “other” adherents provided by the U.S. Religious Census. The “other” adherents category includes groups that are typically the smallest and least mainstream groups in the American religious landscape; they include all groups other than Roman Catholics and Evangelical, Black and Mainline Protestants. Finally, I normalized the variable’s distribution using a natural log transformation to adjust the original distribution’s substantial positive skew. Note that this variable is a transformed indicator of actual population figures, not a proportion of the total population in any locality. I chose this

---

39 Data from the 2010 U.S. Religious Census is available for download from the Association of Religion Data Archives at http://thearda.com/Archive/Files/Descriptions/RCMSMT10.asp.
40 See http://thearda.com/rcms2010/selectMetro.asp to view all metropolitan and micropolitan adherence figures.
41 Members of the Church of Jesus Christ of Latter-day Saints (also known as the Mormons) were also included in the “other” adherents category. However, given their numerical majority status in Salt Lake City, Utah, where claims in that state were heard, as well as their considerable representation in a number of other metropolitan statistical areas (e.g., Boise, Idaho), I did not include Mormons in my minority population totals.
approach on the premise that population totals would be better indicators of judges’ aggregate exposure to members of religious minority faiths in the course of their daily lives.\textsuperscript{42}

To estimate the effect of state precedent for using a balancing test, I generated a dichotomous precedent measure. A case was coded with a “1” on this measure if there was a state court precedent for using this type of test prior to when the case was decided. Cases received this value for one of three reasons:

- the highest court in the state where the case was heard had previously interpreted its state constitution to require the use of a balancing test in adjudicating such cases;

- the state legislature had enacted legislation requiring the use of the balancing test in such cases (i.e., a state version of the federal Religious Freedom Restoration Act [RFRA] of 1993); or

- the case concerned prisoners’ religious rights or a religious land use dispute and was ruled upon after June 22, 2000, the date when the federal Religious Land Use and Institutionalized Persons Act (RLUIPA) came into force requiring the use of a form of balancing test for adjudicating these types of cases.

Determinations about the first criterion were made based on the list of state precedents provided in Durham and Smith (2010:89-90). Determinations on the second criterion were made after I obtained the dates that state-RFRAs came into force through searches of state legislation databases and online editions of major state newspapers.

\textsuperscript{42} As an alternative to this measure of religious diversity, Herfindahl Index (HHI) values were calculated at the county level for each of the areas under consideration. This measure of religious market concentration suggests that a number of counties with high religious minority populations (for instance, the counties where the cities of Boston, MA and Los Angeles, CA are located, with respective Herfindahl Index values of .567 and .487) are far less pluralistic than counties with much smaller religious minority populations (for instance, the counties where Little Rock, AR and Boise, ID are located, with respective Herfindahl Index values of .192 and .201). Thus, religious diversity measures based on market concentration may be a misleading indicator of the magnitude of minority religions’ presence in a given locality.
As an additional control, I accounted for state-level religious salience using data from the Pew Forum’s U.S. Religious Landscape Survey (2008). The survey asked respondents to identify the importance of religion in their lives; respondents could specify whether they felt religion was very important, somewhat important, not too important or not at all important to them. Using respondents’ answers to this question, I calculated the proportion of each state’s population that felt religion was “very important” in their lives.

**Methods of Statistical Analysis**

Chapters 4 through 6 present and describe the results of quantitative analysis based on the data and measures I have discussed in this chapter. Chapter 4 summarizes the univariate structure of the data and describes the distributions of key variables. This approach allows me to discuss a host of descriptive findings, including claimant success rates in free exercise cases and the distribution of claims made by members of different religious groups. I also examine bivariate associations between case outcomes and a host of hypothesized predictor variables; I do so using chi-square-based measures of association such as phi and Cramér’s V as well as results from bivariate binary logistic regression models.

I present and discuss multivariate analyses testing my research hypotheses in Chapters 5 and 6. Chapter 5 focuses on examining factors hypothesized to predict favorable decisions for free exercise claimants. I use logistic regression analysis to estimate a series of models that let me test my hypotheses about this dependent variable. Binary logistic regression methods allow us to estimate the odds of a dichotomous outcome given the values of that outcome’s predictors (Pampel 2000). Analysis in this chapter was conducted using the Stata 12 statistical package.
Chapter 6 focuses in greater depth on the type of legal test judges use in evaluating claims and on the factors that predict whether judges will use a balancing test or a rational-basis test. After examining predictors of balancing test use using binary logistic regression analysis, I estimate structural equation models to account for the simultaneous effects of predictors on legal test use and case outcomes. Analysis in the latter half of the chapter is performed using Mplus, a structural equation modeling program that permits analysis of continuous and categorical outcome variables alike. A particular advantage of Mplus is its capability of allowing users to analyze dichotomous endogenous variables in a structural equation modeling framework, a capability necessary to conduct my analysis in Chapter 6.

While the dataset analyzed in the chapters to come contains a relatively small amount of missing data, it is still important that I account for missingness within cases. This is especially pertinent regarding my measure of religious affiliation. Approximately one claim in five (21 percent) was brought by claimants whose religious affiliation was not provided in the text of the court decisions. However, preliminary logistic regression analysis indicated that claimants with missing affiliation data were not significantly more or less likely to prevail in their claims than those with reported religious affiliations, suggesting that religious affiliations were missing at random (Allison 2002:4).

Operating under this assumption, I estimated all models in Chapter 5 using the MI module in the Stata 12 statistical package. This module permitted the use of multiple imputation under the multivariate normal model to address missingness on both the religious affiliation variable and interaction terms calculated from this variable. I also utilized a multiple imputation procedure in Mplus to address missingness when estimating the structure equation models in Chapter 6. Religious affiliation values were also imputed for claimants who were identified as
“Christian” in the text of decisions but whose affiliations were not further specified. A dichotomous indicator identifying whether or not a claimant was Christian was included to provide additional information in the imputation procedure. Twenty datasets were created using a Markov chain Monte Carlo algorithm, and estimates from these datasets were pooled using Rubin’s (1987) combination procedure to yield the estimates reported in Chapters 5 and 6.
CHAPTER 4: DESCRIPTIVE STATISTICS AND BIVARIATE ASSOCIATIONS

Introduction

Chapter 4 begins my review of the data on state appellate free exercise claims in America from 1997 to 2011. This chapter discusses the univariate distributions of variables discussed in the previous chapter and presents basic bivariate measures of association between case outcomes and their hypothesized predictors. The chapter is structured as follows. First, I review the frequency with which free exercise claims in the state courts were successful during this period. Next, I review the distributions of additional variables, including the type of legal test used in evaluating each claim, claimants’ religious affiliations and the substantive nature of each claims (i.e., case type). I consider the associations between successful free exercise claims as well as balancing test use and a host of independent variables on a bivariate basis, and I conclude by providing a summary of the descriptive findings from this chapter.

Reviewing these variables’ univariate distributions is important for several reasons. First, the univariate distribution of free exercise outcomes I discuss in this chapter allows us to understand how common or rare it is for claimants to prevail in these cases. While the multivariate analyses in subsequent chapters highlight the factors that predict greater or lesser success, they say little about how often claimants prevail. Second, a review of the distribution of legal test use enables us to see how balancing test use has been limited or augmented since the Supreme Court’s decision in City of Boerne v. Flores, which invalidated Congress’ efforts to make its use more common at the state level.\(^{43}\) While Flores appeared to limit the scope of balancing test use in state courts, the descriptive analysis in this chapter allows us to see the extent to which use of this test persists. Finally, the univariate distribution of claims by religious

---

\(^{43}\) 521 U.S. 507 (1997)
affiliation allows us to see whether members of minority faiths are disproportionately likely to make free exercise claims in courts; I compare the distribution of claims with the representation of each group in the total U.S. population. In addition, the bivariate measures of association reviewed in this chapter provide a basic understanding of the nature of relationships between variables and foreshadow results from the hypothesis tests conducted in Chapters 5 and 6. As will be seen throughout this chapter, free exercise claimants have a very difficult time prevailing in the state appellate courts, and it is very unlikely that any particular free exercise claim will be upheld.

**Success and Failure in Free Exercise Claims**

I begin by reviewing the proportion of free exercise claims that were upheld from mid-1997 through the end of 2011. As shown in Figure 1 on the following page, claimants were overwhelmingly unsuccessful in the state appellate courts. During the period from the Supreme Court’s decision in *City of Boerne v. Flores* in June of 1997 to the end of 2011, only 12% of free exercise claims were resolved in favor of the claimant, a total of 26 claims. This stark finding contrasts with previous research suggesting that free exercise claimants tended to be more successful than my analysis shows. For instance, Brent (1999:250) found that 26% of free exercise claims were successful in the U.S. Court of Appeals between 1987 and 1996, while Sisk and his colleagues (2004:555) observed a success rate of 35.6% across all federal courts over roughly the same span of time. Using judicial votes rather than case outcomes as a unit of analysis, Claborn (2011:631) observes an even greater success rate of 45% at the state level between 1998 and 2005.
Several potential factors may account for this discrepancy. First, it is possible that prior studies employed broader definitions of what constitute free exercise claims. As discussed in Chapter 3, the universe of decisions I study is made up of opinions involving free exercise claims made under state or federal constitutional protections as well as those made under free exercise rights-expanding legislation, such as RLUIPA or state Religious Freedom Restoration Acts. To be considered for analysis, claims must have been adjudicated using either a balancing test or a rational-basis test. This excludes disputes over legal benefits that religious groups enjoy but which non-religious groups do not, such as exemptions from taxes or anti-discrimination laws (i.e., the ministerial exception). Previous studies may have included these types of cases in their analysis; indeed, both Brent (1999: 258) and Sisk et al. (2004: 555) explicitly study claims involving tax exemptions for religious groups.

Including these types of cases in the analysis appears to increase the success rate to a level more consistent with prior research. I had initially cast a wide net in my attempts to locate
all state appellate free exercise decisions and had coded an additional 228 opinions from the LoislawConnect legal database beyond the 225 I analyze in these chapters. This additional set of cases included disputes over tax exemptions, ministerial exceptions and other matters not directly litigating the meaning of constitutional free exercise provisions. When all 453 of these decisions are considered, I find that claimants received favorable judgments approximately 33 percent of the time (i.e., in 149 of 453 cases), a figure more consistent with success rates reported in previous social scientific studies. Claims over the denial of religious tax exemptions are particularly successful; claimants received favorable decisions on these matters a remarkable 58 percent of the time (in 48 of 83 total cases). I am inclined to believe that the discrepancy in success rates between the present study and prior social scientific research is due largely to this case selection issue, especially given that legal scholars examining narrowly-defined free exercise disputes observed similarly low success rates. Ryan (1992:1416-1417) observed a 12 percent claimant success rate in federal courts of appeals from 1980 to 1990, while Lupu (1998:590-591) noted a success rate of 15 percent in claims based on RFRA prior to its state-level invalidation in City of Boerne v. Flores. In short, I believe that the low claimant success rate I observe is due predominantly to what I argue is an appropriately narrow focus on free exercise disputes not involving specific and unique benefits religious groups already enjoy.

However, it is also possible that the present study’s focus on claims heard in the state courts may be partially responsible for the lower success rate. As described in Chapter 2, the Supreme Court stated unambiguously in City of Boerne v. Flores (1997) that the state courts could not use the compelling interest test (a form of balancing test) as mandated by the Religious Freedom Restoration Act (RFRA) to adjudicate free exercise claims made under the First Amendment’s Free Exercise Clause. However, the Court was notably vague on whether federal
courts could still rely on RFRA’s compelling interest test when ruling on these claims, and in fact, the Court would hold in 2006 that RFRA did still control free exercise claims in the federal courts.\footnote{Gonzales v. O Centro Espirita Beneficente Uniao do Vegetal, 546 U.S. 418 (2006)} While this clear line of separation between federal and states claims was drawn almost a decade after \textit{Flores}, savvy litigants may have reasoned that their claims stood a better chance of success in a federal court system marked by uncertainty than in the state courts. While it is not possible to compare state and federal success rates in the present study, it is conceivable that claims in the federal courts after 1997 were more likely to be adjudicated using a balancing test than those in the state courts, and thus there may be a higher overall success rate in the federal courts than at the state level.

Regardless of these alternative explanations, the fact remains that it is remarkably difficult for claimants to prevail in these cases. To start to understand why, I now review the univariate descriptions of several key independent variables which may influence case outcomes, as well as to the bivariate associations between these variables and the outcomes of claims in the state courts.

\textbf{Legal Tests and Successful Free Exercise Claims}

Figure 2 on the following page shows the distribution of legal tests used to adjudicate the 225 free exercise claims heard in state appellate courts between mid-1997 and 2011. This figure demonstrates that despite the Court’s holding in \textit{City of Boerne v. Flores}, the use of the rational-basis test did not dominate free exercise jurisprudence in the states at the end of the 20\textsuperscript{th} and beginning of the 21\textsuperscript{st} century. On the contrary, actions by state judges and lawmakers at the state and federal levels appear to have countered the Supreme Court’s decision substantially, creating an environment in which more than half of all state free exercise claims were decided using some
form of balancing test. Some 56 percent of claims were adjudicated using a form of balancing test in the state courts, be it the compelling interest test articulated in *Sherbert v. Verner* or a less protective balancing test. The remaining 44% of cases were adjudicated using a rational-basis test, most commonly the test articulated by the Supreme Court in *Employment Division v. Smith*.

In their review of the state of free exercise rights in America, Durham and Smith (2010) concluded that the legal opportunity structure at the state level had grown more favorable to claimants in the wake of *City of Boerne v. Flores*. Their conclusion was based on the factors I described in Chapter 2: diffusion of state legislation requiring the use of the compelling interest test (a form of balancing test), the passage of protective federal legislation in the form of 2000’s Religious Land Use and Institutionalized Persons Act (RLUIPA) and the establishment of judicial precedents requiring a balancing test when claims based on state-level constitutional free exercise provisions are evaluated. Taken together, they argued, these processes suggested that claims in the states might not necessarily be evaluated using the more restrictive rational-basis
Table 3: Case Outcomes and Legal Tests in State Appellate Free Exercise Decisions

<table>
<thead>
<tr>
<th>Outcome of Decision</th>
<th>Rational-Basis Test</th>
<th>Balancing Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable Decision</td>
<td>5 (5.05%)</td>
<td>21 (16.67%)</td>
</tr>
<tr>
<td>Unfavorable Decision</td>
<td>94 (94.95%)</td>
<td>105 (83.33%)</td>
</tr>
<tr>
<td>Total</td>
<td>99 (100%)</td>
<td>126 (100%)</td>
</tr>
</tbody>
</table>

test, as the Court’s decision in *City of Boerne v. Flores* might suggest. The evidence from my analysis of state court opinions supports their conclusion; claimants were somewhat more likely to see their claims evaluated according to a balancing test than a rational-basis test in *Flores’* wake.

A review of the bivariate association between legal test use and case outcomes, as depicted in Table 3 above, suggests that even though balancing test use is more prevalent in the states, free exercise claimants are still at a decided disadvantage in court. A mere 17 percent of claims evaluated using a balancing test resulted in favorable decisions for claimants. This success rate is significantly different than the success rate among claims evaluated using a balancing test ($p < .01$), in which approximately five percent of cases were resolved in favor of the claimant ($\chi^2 = 7.32$, df = 1). The phi statistic ($\phi = .18$), a chi-square-based statistic that evaluates the fourfold point correlation between successful claims and the type of legal test used to evaluate them, indicates a moderate association between these variables (Healey 2005:342). Despite the existence of this association, however, the low overall success rates in cases evaluated under both types of tests indicate that free exercise claimants are at a substantial disadvantage in the state appellate courts.
Religious Affiliation and Successful Free Exercise Claims

Claimants’ religious affiliations, as noted in Chapter 3, were recorded from content analysis of court decisions and were initially coded into one of fourteen religious affiliation categories. These categories were largely the same as those used in the U.S. Religious Landscape Survey conducted by the Pew Forum on Religion & Public Life (2008), though denominational Protestants were classified into liberal, moderate, conservative and sect categories based on Iannaccone’s (1994) strictness typology. Figure 3 on the following page describes the percentage of all state free exercise claims made by different religious groups during the period of analysis (collapsed into seven religious categories), and compares it to the total representation of that group in the U.S. population circa 2007 according to the Pew Forum’s U.S. Religious Landscape Survey (2008).

Consistent with past research (e.g., Wybraniec and Finke 2001), Figure 3 shows that the corpus of free exercise claims are not made proportionately across religious groups in America. This is especially true among members of some of the groups that are the furthest from the American religious mainstream, namely new religious movements (NRMs) and members of Native American religious groups. Taken together, these groups comprise approximately one percent of the United States religious population, yet the groups made 10 percent of all free exercise claims in the state appellate since mid-1997. As discussed in Chapter 1, we should expect to see that groups furthest outside of the mainstream will shoulder the heaviest burdens on their religious practice, and Figure 3 suggests that this is the case. A similar pattern may be seen among members of Protestant sects, which in this classification include Pentecostals, Jehovah’s Witnesses, members of the Church of the Nazarene and other members of the strictest Protestant denominations. Members of these groups comprise about nine percent of the American religious affiliation.

Atheists, agnostics and those who list no particular religious affiliation are not included in Figure 3.
landscape, yet they account for approximately 15 percent of all state appellate free exercise claims.

Conversely, members of more mainstream faiths account for disproportionately low numbers of claims. Christians who are not classified as Catholics or members of Protestant sects account for nearly three-fifths of all religious believers in America (58 percent), yet only make approximately two-fifths of all free exercise claims in state appellate court (41 percent). Figure 3 demonstrates a similar pattern among Roman Catholics; despite comprising 29 percent of the U.S. religious population, Catholics were the claimants in just 21 percent of cases.

A closer look at claims made by denominational Protestants demonstrates how religious marginality influences the likelihood of making claims. Figure 4 below compares the
percentages of free exercise claims made by groups of denominational Protestants (outside of the historically Black Protestant tradition) to each groups’ shares of the American Protestant denominational population in 2007. Members of liberal and moderate Protestant denominations as well as Protestant sects each account for about one-fifth of the total share of denominational Protestants in America (23 percent, 18 percent and 19 percent, respectively), while conservative Protestants account for the remaining two-fifths (41 percent). As shown in Figure 4, however, the number of claims made across these groups increases exponentially with their level of organizational strictness. Members of Protestant sects made an overwhelming majority (68 percent) of all denominational Protestant free exercise claims from 1997 to 2011, a far greater share than would be expected if claims had been made proportionally across groups. Conversely, members of more mainstream liberal, moderate and conservative Protestant

---

**Figure 4: Distribution of Free Exercise Claims Made By Denominational Protestants**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberal</td>
<td>4%</td>
<td>23%</td>
</tr>
<tr>
<td>Moderate</td>
<td>9%</td>
<td>18%</td>
</tr>
<tr>
<td>Conservative</td>
<td>19%</td>
<td>19%</td>
</tr>
<tr>
<td>Sect</td>
<td>68%</td>
<td>19%</td>
</tr>
</tbody>
</table>

46 The latter estimates are derived from the Pew Forum’s U.S. Religious Landscape Survey (2008). Note that the claim percentage figures shown here are in comparison to to denominational Protestants, not to all religious groups.
denominations accounted for less than one-third of all state free exercise claims. This discrepancy is most striking for members of liberal Protestant groups; such groups account for 23 percent of all denominational Protestants, but liberal Protestants made just four percent of all denominational Protestant free exercise claims.

Taken together, Figures 3 and 4 suggest that secular laws and government policy are far more likely to burden minority religious faiths than mainstream faiths, if the number of free exercise claims made in the state courts is any indication. Faiths outside of the American religious majority are more likely to ask for assistance from the courts in disputes over their ability to practice their religion freely. This is consistent with Black’s (1976:114) argument that the law tends to favor majorities at the expense of societal outsiders.

Once disputes reach the courts, however, are some groups more likely to prevail than others? Figure 5 on the next page suggests that there is little pattern in who wins or loses: no religious group is especially likely to prevail. Indeed, the chi-square measure of the bivariate association between success in court and the claimants’ religious affiliation using this categorization scheme is not statistically significant ($\chi^2 = 4.31, \text{df} = 6$). Christians (excluding Catholics and members of Protestant sects) do enjoy a somewhat higher success rate of 15% (10 wins in 67 claims) than most other groups, while Catholics (7% success), members of new religious movements and Native American faiths (8% success), and Jews (0% success) seldom ever win. The small number of claims brought by some of these groups discourages me from generalizing about group outcomes based on Figure 5, but one conclusion that we can certainly draw is that the odds of success are stacked against claimants no matter their religious affiliation.

47 The Cramér’s $V$ statistic for the association, which is based on the chi-square statistic, has a value of .15, which suggests a moderate if small association between these variables; nevertheless, the chi-square statistic assessing this association is not statistically significant itself.
An alternate way to assess the bivariate association between case outcomes and claimants’ religious affiliation – an approach that I employ in the multivariate analysis in subsequent chapters – is to consider the size of the religious group in the population. As described in Chapter 3, I coded claimants’ religious affiliations into one of fourteen categories and then calculated the state population proportion that belongs to the same religious category using U.S. Religious Landscape Survey (2008) data. I then examined whether there is an association between success in court and the proportion of the state population who belong to the same religious categorization as the claimant. The biserial correlation between these measures is not statistically significant ($\rho = -.05$, S.E. = .101), though it is worth noting that the negative

---

48 As noted in Chapter 3, I used an adjusted logit transformation so that this variable’s distribution would more closely approximate normality. The biserial correlation described here is between favorable decisions in cases and logit-transformed local religious populations in imputed dataset $m = 1$.  

86
value of this coefficient indicates that religious groups with greater representation in their states’ populations fare worse than groups with less representation. This may be due in part to the low success rate among Catholics, who are routinely well-represented in state populations. This may also be due in part to the comparatively high success rate among Muslims, who are far less prevalent in their states’ populations but were successful in nearly 24 percent of their claims. Regardless of how much each religious group might influence the direction of this association, the association itself is not statistically significant.

Case Types and Successful Free Exercise Claims

In addition to the type of legal test used to evaluate claims and the religious affiliation of the claimants, I coded the substantive nature of each free exercise claim. Table 4 on the following page provides the frequency distribution of the fourteen types of cases based on the coding instrument as well as two additional case types identified from among the category of “Other” case types. For analytic purposes, I group these case types into four categories: prison and land use rights disputes, disputes over positive government policies (i.e., disputes over actions that government actors require all citizens to take), disputes over education and child welfare, and other types of disputes.

Perhaps the most striking feature of Table 4 is how common the two types of substantive disputes that are to be adjudicated using the compelling interest test (a form of balancing test) under the Religious Land Use and Institutionalized Persons Act (RLUIPA) are in the case distribution. Claims involving prisoners’ religious rights and religious land use disputes under RLUIPA – prison rights and land use disputes – are found in 81 opinions, or more than one-third of all free exercise cases adjudicated at the state appellate level. Whether this is because the use
Table 4: Distribution of Claims by Substantive Nature of Dispute

<table>
<thead>
<tr>
<th>Nature of Dispute</th>
<th>Frequency (Percentage of Claims)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prison/Land Use Disputes (36.00% of claims)</strong></td>
<td></td>
</tr>
<tr>
<td>Prison Rights Dispute</td>
<td>33 (14.67%)</td>
</tr>
<tr>
<td>Land Use Rights Dispute</td>
<td>48 (21.33%)</td>
</tr>
<tr>
<td><strong>Disputes Over Positive Government Policies (14.67% of claims)</strong></td>
<td></td>
</tr>
<tr>
<td>Immunization Dispute</td>
<td>2 (0.89%)</td>
</tr>
<tr>
<td>Government Identification Dispute</td>
<td>15 (6.67%)</td>
</tr>
<tr>
<td>Other Government Operations Dispute</td>
<td>16 (7.11%)</td>
</tr>
<tr>
<td><strong>Education and Child Welfare Disputes (19.56% of claims)</strong></td>
<td></td>
</tr>
<tr>
<td>Education Dispute</td>
<td>16 (7.11%)</td>
</tr>
<tr>
<td>Other Child Welfare Dispute</td>
<td>28 (12.44%)</td>
</tr>
<tr>
<td><strong>Other Disputes (29.78% of claims)</strong></td>
<td></td>
</tr>
<tr>
<td>Drug/Banned Substance Dispute</td>
<td>16 (7.11%)</td>
</tr>
<tr>
<td>Dispute Over Other Criminal Violation</td>
<td>14 (6.22%)</td>
</tr>
<tr>
<td>Unemployment Compensation Dispute</td>
<td>6 (2.67%)</td>
</tr>
<tr>
<td>Other Employment-Related Dispute</td>
<td>1 (0.44%)</td>
</tr>
<tr>
<td>Tax Dispute (evaluated using balancing or rational-basis test)</td>
<td>2 (0.89%)</td>
</tr>
<tr>
<td>Dispute Over Internal Operation of Religious Organization</td>
<td>1 (0.44%)</td>
</tr>
<tr>
<td>Other Dispute: Exemption from Tort Liability</td>
<td>7 (3.11%)</td>
</tr>
<tr>
<td>Other Dispute: Claim That Law is Unconstitutional</td>
<td>9 (4.00%)</td>
</tr>
<tr>
<td>Other Dispute, Not Elsewhere Classified</td>
<td>11 (4.89%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>225 (100.00%)</td>
</tr>
</tbody>
</table>
of a balancing test is mandated by law in these cases (a factor which may have given claimants greater confidence that they would receive favorable decisions), or whether these types of claims would have comprised such a large share of the total claims absent these considerations is unclear.

Do different types of cases more prone to success in court than others? Figure 6 above compares the number of claims made on each of the four main case type categories with the number of favorable decisions claimants received.\(^49\) This figure highlights once again how rare it is for claimants to receive a favorable decision; success rates for all cases types are low. As expected, claims eligible to be adjudicated under RLUIPA are somewhat more likely to be upheld (a 21 percent success rate) than other types of claims (the nine percent success rate for “other” cases is the highest success rate among the other three categories). Yet on a bivariate

\(^{49}\) Because the utility of case type as a predictor of successful free exercise claims is most relevant after the passage of RLUIPA in September 2000, Figure 6 only includes those claims that were ruled upon after this date (i.e., those from September 22, 2000 to December 31, 2011).
basis, the nature of claims and the direction of their outcomes are not significantly associated ($\chi^2 = 5.81$, df = 3). The Cramér’s $V$ statistic indicates that there is a moderate association between these variables ($V = .18$), yet the chi-square test fails to observe that this association is statistically significant on a bivariate basis.

**Bivariate Associations and Chapter Summary**

Finally, in order to provide a preliminary description of associations between my two dependent variables (balancing test use and case outcomes) and hypothesized predictors, I regressed both of the dependent variables on the predictors individually. Table 5 on the following page reports the results of each of these logistic regression models. Each of the coefficients in this table (expressed as an odds ratio) is the result of a single bivariate logistic regression model, controlling for no other predictors. Some preliminary conclusions about the associations between variables can be described here. First, as hypothesized, there is a strong, positive bivariate association between judges’ use of a balancing test and the likelihood that cases will be decided in favor of the claimants. The odds of favorable outcomes for claimants are 276 percent greater when judges use a balancing test to evaluate their claims than when they use a rational-basis test. However, this is the only association between hypothesized predictors and favorable decisions that is statistically significant with greater than 95 percent confidence on a bivariate basis. There is one bivariate effect that is marginally significant ($p < .10$), which

---

50 I regressed the dependent variables on the dichotomous case type indicators together and did the same with the judicial retention system indicators. The odds ratios describing the relationships between case type and the dependent variables only refer to the relationships between these variables after RLUIPA was enacted in September 2000.

51 For more information on the bivariate associations I describe here, see Appendix B, which compares frequencies on nominal or dichotomous variables given (a) different case outcomes and (b) the use of one form of legal test or the other. Appendix B also provides the means of continuous variables given different case outcomes and use of different legal tests.
Table 5: Logistic Regression Estimates of Bivariate Associations, 1997-2011

<table>
<thead>
<tr>
<th></th>
<th>Use of Balancing Test</th>
<th>Favorable Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of Balancing Test</td>
<td></td>
<td>3.760**</td>
</tr>
<tr>
<td>Religious Group Size, Claimant</td>
<td></td>
<td>.110</td>
</tr>
<tr>
<td>(Proportion of State Population, Logit)</td>
<td></td>
<td>.897</td>
</tr>
<tr>
<td>Case Type (post-RLUIPA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/Child Welfare Dispute</td>
<td>.790</td>
<td>.479</td>
</tr>
<tr>
<td>Disputes over Positive Gov’t Policies</td>
<td>1.692</td>
<td>.479</td>
</tr>
<tr>
<td>Other Disputes</td>
<td>.673</td>
<td>.386†</td>
</tr>
<tr>
<td>Metro Area Minority Pop Size (Log)</td>
<td>1.051</td>
<td>1.081</td>
</tr>
<tr>
<td>Proportion of GOP Judges</td>
<td>.638</td>
<td>1.929</td>
</tr>
<tr>
<td>Judicial Retention System</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>1.220</td>
<td>.633</td>
</tr>
<tr>
<td>Competitive Election</td>
<td>1.434</td>
<td>1.224</td>
</tr>
<tr>
<td>Statewide Importance of Religion</td>
<td>.045*</td>
<td>.137</td>
</tr>
<tr>
<td>Precedent</td>
<td>1.882*</td>
<td>****</td>
</tr>
</tbody>
</table>

*** p<.001; ** p<.01; * p<.05; † p<.10. N = 225 using 20 imputations. Binary logistic regression coefficients expressed as logged odds.

Case Type: Prisoners’ Rights/Land Use Disputes
Judicial Retention System: No Direct Election.

Concerns the substantive nature of claims; prisoners’ rights and land use claims are more likely to be upheld than cases in the omnibus “other disputes” category, which includes most criminal matters and unemployment compensation claims. However, none of the other associations shown in this table are statistically significant; religious group size is not associated with differences in case outcomes, nor is the political composition of the judicial panel, the local religious minority population size or the state’s judicial retention system.
Few of the hypothesized associations between balancing test use and its predictors are statistically significant on a bivariate basis either. As hypothesized, use of a balancing test is significantly more likely when there is state precedent for its use. Curiously, there is a negative association between statewide religious salience and the use of the balancing test, but no other hypothesized predictor is associated with its use. I test my hypotheses in a more formal manner in the following two chapters in a multivariate logistic regression framework.

This chapter presented an overview of the univariate distributions of variables that are central to my analysis and assessed basic bivariate associations between these variables. The primary purpose of the analysis in this chapter was to answer descriptive questions about the processes related to the adjudication of religious free exercise cases in the state courts. Several findings from this chapter are worth emphasizing.

First, the distribution of case outcomes revealed the difficulty claimants experience in pursuing their free exercise claims in court; only 12 percent of all state appellate claims were upheld. This claimant success rate is substantially lower than had been observed in other studies of free exercise disputes by sociologists and political scientists, but was consistent with the more targeted analyses of legal scholars who explored free exercise success rates. Second, the distribution of legal test usage since 1997 was divided nearly in half between balancing tests and rational-basis tests. This finding is remarkable given that the Supreme Court mandated the use of a rational-basis test at the state level for evaluating claims based on the First Amendment’s Free Exercise Clause in its 1997 City of Boerne v. Flores decision. The combination of state and federal legislative interventions since then, along with robust libertarian judicial interpretations of state constitutional free exercise clauses, has broadened claimants’ access to balancing tests in the courts. Moreover, evidence from this chapter indicates that the use of balancing tests by
judges in evaluating free exercise claims is associated with higher likelihoods of claimant successes. Finally, this chapter provided evidence that while religious minorities’ ability to practice their religions freely are disproportionately limited by the law and government policy, there is little difference in outcomes of free exercise cases between religious groups in the state appellate courts: all groups are at a marked disadvantage vis-à-vis the state.

The main lesson of this analysis is that free exercise claimants are overwhelmingly unsuccessful in the state appellate courts. While there are some variables that predict successful claims on a bivariate basis (particularly the form of legal test judges use to evaluate claims), the bivariate analysis reported in this chapter yielded very few significant associations between case outcomes and any independent variables. This is likely due in part to the small number of successful claims overall. Yet until additional factors are controlled for simultaneously, it would be premature to draw firm conclusions about the nature of the relationships between these variables. Chapters 5 and 6 explore the relationships introduced in this chapter using multivariate statistical analysis, and the hypotheses laid out in Chapter 2 are tested formally in these chapters.
CHAPTER 5: MULTIVARIATE ANALYSIS OF CASE OUTCOMES

Introduction

What factors influence whether state appellate courts will uphold or reject religious free exercise claims? Hypotheses based on legal positivist and legal realist logics suggest alternate explanations for how courts reach different outcomes. Positivist hypotheses suggest that variations in case outcomes are the result of neutral application of the law. Realist hypotheses, on the other hand, suggest that judges’ attitudes and desire to maximize their self-interest influence outcomes as well. The descriptive analysis presented in Chapter 4 highlighted the difficulty claimants experience in the courts; free exercise claim success rates in the state appellate courts have been very low since 1997. Nevertheless, the bivariate findings reported in the previous chapter suggest that some factors increase the likelihood of claimant success; cases decided using a legal test that accounts for the severity of burdens on claimants’ free exercise (i.e., a balancing test) were more likely to be resolved favorably for claimants without controlling for other factors, and certain types of cases (namely those involving prisoners’ rights and land use disputes) were more somewhat likely to yield claimant successes than other cases.

Chapter 5 uses multivariate statistical analysis to examine these hypothesized relationships more closely, controlling for other factors that might be related to differences in case outcomes. In other words, the analysis in this chapter as well as in Chapter 6 allows me to account for the potential that the bivariate associations observed in the previous chapter might be due to the spurious effects of other variables. Like the descriptive analysis summarized in Chapter 4, the cases analyzed here were ruled upon in state intermediate appellate courts and state courts of last resort from June 25, 1997 through December 31, 2011 unless otherwise noted.
This chapter and the next one present analyses I conducted to answer the two research questions at the heart of this dissertation: (1) what are the factors that predict whether a free exercise claim will be upheld or rejected in court, and (2) what are the factors that predict which of the two legal tests (i.e., a rational-basis or balancing test) judges will be use in adjudicating a free exercise claim? The primary focus of Chapter 6 is on the latter question; it explores the factors that predict the type of legal test that is used. The present chapter deals explicitly with the former question. In the following analysis, I present and discuss the results of binary logistic regression models examining factors that predict whether claimants will receive favorable decisions when pursuing their claims in the state appellate courts.

I test six hypotheses in this chapter, three based on the logic of legal positivism and three based on legal realism. Hypotheses of each type are first tested together as a set and are then tested in fully-specified models. The hypotheses I test in this chapter that are informed by a positivist logic are:

- Claimants should be more likely to receive favorable decisions when courts adjudicate their claims using a balancing test instead of a rational-basis test (Hypothesis One);
- Claims made by members of religious minority groups should be less likely to be upheld than claims made by members of more mainstream faiths (Hypothesis Two); and
- After the year 2000 (due to Congress enacting the Religious Land Use and Institutionalized Persons Act), claims involving prisoners’ religious rights and disputes over religious land use should be more likely to be upheld than claims over other substantive issues (Hypothesis Three-B).

---

52 Hypothesis Four is not tested in this chapter because it posits legal test use as the dependent variable.
A number of studies attest to the predictive power of “legal period” on case outcomes; in other words, the year a case was decided impacts the likelihood of it being resolved in favor of the claimant (Wybraniec and Finke 2001; Brent 1999, 2003). This finding is likely due to the prevalence of different forms of legal tests used for evaluating claims during the separate legal periods. Accounting directly for the form of legal test judges use should substantially explain differences in outcomes. Likewise, past research has found that members of minority faiths are less likely to prevail in their claims than are members of more mainstream faiths (Brent 1999; Wybraniec and Finke 2001; Adamczyk et al. 2004). This may be due less to judges’ biases against members of minority faiths and more to the likelihood that minorities’ religious behavior will be outside of what the law permits.

In contrast, realist hypotheses are based on the proposition that judges are rational actors who wish to maximize their own interests. These hypotheses emphasize the role of judicial preferences and structural constraints in decision-making. Consistent with this approach to understanding judicial behavior, I test the following hypotheses:

- Claimants should be less likely to receive favorable decisions as the Republican composition of the judicial panel hearing claims increases (Hypothesis Five-B);
- In states where judges do not require public approval to retain their seats (e.g., through direct election), claims raised by members of minority faiths should be more likely to be upheld than in states where judges face popular referenda (Hypothesis Six-B); and
- Free exercise claims will be more likely to be resolved in favor of the claimant as the population of religious minorities in the localities where the claims are heard increases (Hypothesis Seven-B).
Each of these hypotheses is suggested by the literature on free exercise disputes and judicial behavior more generally. Previous research finds that Republican-affiliated judges are less likely to uphold civil liberties claims (Segal and Cover 1989; Marshall and Ignagni 1994); I hypothesize that this effect will persist in studies of claims over religious free exercise rights. Likewise, studies suggest that judges are likely to act in ways that benefit their own self-interest (Cohen 1991; Epstein et al. 2013). Much of the literature exploring rational-choice explanations for judicial behavior has focused on the federal judiciary (Posner 1993; Schauer 2000; Epstein et al. 2013), but the present study allows me to apply insights from this literature to a study of the state courts. For instance, judges in many states face election and re-election; I hypothesize that judges who do so will be less inclined to support claims brought by potentially unpopular religious minorities than judges who do not face direct public scrutiny for their actions. Finally, consistent with the findings of Sisk et al. (2004) and the argument that heightened exposure to religious minorities will influence judges to be more sympathetic to burdens imposed on religious exercise, I hypothesize that local minority population size and the likelihood of favorable decisions for claimants should be positively associated.

Chapter 5 is organized as follows. In the first section, I report findings from across the full temporal window of analysis. This allows me to test all of my hypotheses regarding the outcomes of free exercise decisions (with the exception of my hypothesis about case type, though I still control for this variable), and results from these tests are reported in this first section. In the second section, I explore the relationship between favorable decisions and one particular independent variable: claimants’ religious affiliations. In particular, I explore whether alternate conceptualizations of religious affiliation have an effect on the relationship between the outcomes of cases and claimants’ religious backgrounds. In the final section, I present an
additional set of models to test my hypothesis about whether the substantive nature of cases (i.e., “type” of case) has an impact on case outcomes. These models are discussed separately because the hypothesis is only valid for cases heard after September of 2000, when the Religious Land Use and Institutionalized Persons Act (RLUIPA) was signed into law. RLUIPA required that particular types of free exercise disputes (those involving prisoners’ rights and land use disputes) be adjudicated under the compelling interest test, a form of a balancing test. I hypothesize that a difference in the outcomes of these types of cases is due to the difference in the form of legal test judges use, but this is only valid after RLUIPA came into force. I conclude the chapter with a brief summary of findings from these models.

**Predicting the Outcomes of Free Exercise Cases**

First, I estimated a series of binary logistic regression models to predict the likelihood that a free exercise claim would be upheld (i.e., decided in favor of the claimant) given the values of various predictor variables. The models are the result of pooling estimates from twenty imputed datasets using the combination procedure developed by Rubin (1987). These models are shown in Table 6 on the following page. Model 1 tests the relationships described in my legal positivist hypotheses, while Model 2 tests legal realist hypotheses. Model 3 controls for the main effects of all factors, while Model 4 tests the interactions between state judicial retention systems and the proportion of the state population classified in the same religious affiliation category as the claimant (Hypothesis Six-B), controlling for all other factors. The coefficients given in Table 6 and in all other tables in this chapter are expressed as odds ratios for
Table 6: Modeling the Likelihood of Successful Free Exercise Claims, 1997-2011

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use of Balancing Test</strong></td>
<td>4.039**</td>
<td>4.005*</td>
<td>4.234*</td>
<td></td>
</tr>
<tr>
<td>Religious Group Size, Claimant</td>
<td>.874</td>
<td>.867</td>
<td>.641†</td>
<td></td>
</tr>
<tr>
<td>(Proportion of State Population, Logit)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Case Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/Child Welfare Disputes</td>
<td>.466</td>
<td>.465</td>
<td>.451</td>
<td></td>
</tr>
<tr>
<td>Disputes over Positive Gov’t. Policies</td>
<td>.423</td>
<td>.398</td>
<td>.455</td>
<td></td>
</tr>
<tr>
<td>Other Disputes</td>
<td>.394†</td>
<td>.411</td>
<td>.446</td>
<td></td>
</tr>
<tr>
<td><strong>Metro Area Minority Pop Size (Log)</strong></td>
<td></td>
<td>1.085</td>
<td>1.086</td>
<td>1.097</td>
</tr>
<tr>
<td><strong>Proportion of GOP Judges</strong></td>
<td></td>
<td>1.873</td>
<td>2.153</td>
<td>2.013</td>
</tr>
<tr>
<td><strong>Judicial Retention System</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>.673</td>
<td>.788</td>
<td>2.651</td>
<td></td>
</tr>
<tr>
<td>Competitive Election</td>
<td>1.491</td>
<td>1.947</td>
<td>6.401</td>
<td></td>
</tr>
<tr>
<td><strong>Statewide Importance of Religion</strong></td>
<td></td>
<td>.073</td>
<td>.062</td>
<td></td>
</tr>
<tr>
<td><strong>Interaction: Judicial Selection x Religious Group Size</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election x Religious Group Size</td>
<td></td>
<td></td>
<td>1.604†</td>
<td></td>
</tr>
<tr>
<td>Competitive Election x Religious Group Size</td>
<td></td>
<td></td>
<td>1.518</td>
<td></td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>.013***</td>
<td>.040**</td>
<td>.013*</td>
<td>.005*</td>
</tr>
<tr>
<td>McFadden Pseudo-$R^2$ (mean of 20 imputations)</td>
<td>.081</td>
<td>.026</td>
<td>.114</td>
<td>.145</td>
</tr>
</tbody>
</table>

*** p<.001; ** p<.01; * p<.05; † p<.10.  N = 225 using 20 imputations. Binary logistic regression coefficients expressed as odds ratios.

Reference Categories: **Case Type:** Prisoners’ Rights/Land Use Disputes. **Judicial Retention System:** No Direct Election. **Judicial Retention x Religious Group Size:** No Direct Election x Religious Group Size.
more straightforward interpretation.\footnote{Calculation of changes in the odds of a successful claim given a one-unit change in the value of a predictor variable is done by subtracting 1 from the odds ratios reported in these tables and multiplying the difference by 100 (Pampel 2000:23).}

The broad conclusion to be drawn from the models in Table 6 is that few variables are significantly associated with differences in case outcomes. In general, the models reported in Table 6 provide very limited support for any of the hypotheses they test. Nevertheless, this table does report a major finding. Consistent with Hypothesis One, when judges use a balancing test to evaluate free exercise claims, they are more likely to rule in favor of claimants than when they use a rational-basis test. When controlling for other relationships posited by other positivist hypotheses in Model 1 as well as all other factors in Model 4, claimants are significantly more likely to prevail in court when judges use some form of balancing test to evaluate their claim ($p < .05$). Model 1 indicates the odds of a favorable outcome for claimants were approximately 304 percent greater when a balancing test was used versus a rational-basis test.\footnote{This percentage was calculated as follows: $(4.039 - 1) \times 100 = 303.9$.} The difference in odds of success increased to approximately 323 percent in the fully specified Model 4, controlling for all other factors.

My hypothesis about legal test use was supported more strongly than any of my other five hypothesis concerning case outcomes; this relationship was the only one found to be statistically significant with greater than 95 percent confidence in any of the models estimated for this chapter. This is true even though the models in Table 6 and elsewhere in this chapter yielded both very large and very small odds ratios, both of which would typically indicate significant relationships via logistic regression analyses. Many of these regression coefficients have large standard errors, which are likely due to the low level of variation on the dependent variable (recall that only approximately 12 percent of cases were decided favorably for the
claimant) and to the number of observations included in the analysis (N = 225).\textsuperscript{55} Given these factors, the strength and statistical significance of the relationship between case outcomes and the choice of legal test are especially noteworthy. Moreover, this suggests that policymakers and activists who advocate for legislation expanding the application of balancing tests in these cases are correct in realizing that such legislation broadens the scope of acceptable religious activity.

The results of testing the other two positivist hypotheses will be discussed in subsequent sections. However, the models in Table 6 allow for tests of each of the three realist hypotheses, and it is to these hypotheses that I now turn. As stated in Hypothesis Five-B, claimant successes should be less likely as the proportion of judges affiliated with the Republican Party that hears their claims increases. This hypothesis is not supported by the models in Table 6; neither the partial Model 2, which tests realist hypotheses only, nor the fully specified Model 4 provides evidence for this relationship. In fact, though the relationship is not statistically significant, these models suggest that the opposite may be true; the coefficient for this relationship is positive, suggesting that claimants may fare better when Republicans hold a larger share of seats on the bench. We cannot have even 90 percent confidence that this relationship is true, however, so this suggestion is merely speculative.

Likewise, Hypothesis Seven-B is not supported by the models in Table 6. There is no significant association between the size of the religious minority population (conceptualized here as the population that is neither Protestant, Catholic nor Mormon) in the area where the case was decided and the outcome of the case. The “area” here is defined as the relevant statistical area

\textsuperscript{55} Note that statistical significance need not be of paramount importance in this study, given that the number of cases analyzed represents the apparent universe of published state appellate free exercise decisions, not a sample of such decisions. I would hesitate to generalize these results to the total population of unpublished decisions of this type, since there may be qualitative differences between the two populations; if anything, there may be an even lower claimant success rate in these cases. Nevertheless, given the possibility that some published decision were not included in LoislawConnect at random, or that a random distribution of relevant decisions were not located using my search term, I continue to rely on the concept of statistical significance in interpreting my findings.
containing the city where the case was heard, and is either a metropolitan or micropolitan statistical area depending on the U.S. Census Bureau’s designation. The odds ratios depicting this relationship have values greater than one in both Model 2 and Model 4, indicating that the estimated relationship is in a direction consistent with Hypothesis Seven-B. Despite this, the relationship between religious minority population size and the likelihood of receiving a favorable decision is not statistically significance with even 90 percent confidence.

Yet the legal realist hypotheses were not entirely unsupported by the logistic regression models in Table 6. Hypothesis Six-B, which states that claims involving religious minorities will be more successful in states where judges do not face direct, popular election to retain their seats, was partially supported by the estimates in Model 4. This hypothesis is tested using the interaction terms at the bottom of Table 6; the logit-transformed proportion of a state’s population that belongs to the same religious classification as the claimant was multiplied by a dichotomous indicator of the type of judicial retention system operating where the case was heard. The results in Model 4 indicate the effect on the likelihood of success based on the interaction between religious group size and the use of retention elections is marginally significant \( (p < .10) \), compared with the interaction between religious group size and no direct election.

The nature of this relationship is shown in Figure 7, which plots predicted probabilities of favorable decisions on the Y-axis. This figure was generated using the vibl suite of logistic regression visualization tools for Stata. The tools in this suite deal with the problem of how to represent predicted logistic regression probabilities in a two-dimensional space by reducing the contribution of all other covariates to a single term, expressed as the “linear combination of the remaining predictors in the model multiplied by their corresponding logit coefficient” (Mitchell
Figure 7: Predicting Favorable Decisions by the Interaction of Religious Group Size and Use of Judicial Retention Elections

Values along the X-axis are the values of the logit-transformed religious group size, and values along the Y-axis represent the predicted probabilities of favorable decisions when judges do not face direct election (the broken line) and when they face retention elections (the solid line). The two graphs represent the 20th and 80th percentiles of values of the total contributions of other covariates, which largely explains the marked difference in predicted probabilities given the presence of retention elections in states between the two graphs.
The critical feature in interpreting the interaction is the slope of each line. The figure illustrates that religious minorities fare somewhat better than groups of a larger size in states where appellate judges are not retained through any sort of direct election; this is reflected in the slight negative slope of the dotted line in both graphs. In contrast, the positive slope of the solid lines indicates that when judges are subject to non-competitive retention elections in order to hold their seats, they are more likely to rule in favor of claimants belonging to larger religious groups than they rule for members of smaller groups. When facing popular referenda in order to retain their seats, it appears that judges are less inclined to rule in favor of smaller religious groups than they are for groups of greater size in free exercise disputes. However, a similar effect was not observed for claims evaluated in states where judges face competitive elections against other judicial candidates to retain their seats; the direction of the association remained the same, yet the association was not found to be statistically significant in Model 4.

In all, the positivist hypotheses tend to be supported more strongly than the realist hypotheses. This is reflected in the differences between the mean McFadden pseudo-$R^2$ values given for Model 1 and Model 2. This statistic measures the proportional reduction in error variance in terms of values of the likelihood functions of each model (McFadden 1974). It is not possible to obtain a single pseudo-$R^2$ estimate measuring model fit across the 20 datasets analyzed in this table, but pseudo-$R^2$ values can be obtained using each of these datasets. Model 1, which tests the positivist hypotheses, yielded a mean McFadden pseudo-$R^2$ value of .081 across the 20 datasets (S.D. = .003, with a range from .077 to .088). By comparison, Model 2, which tests the realist hypotheses, yielded a mean value of .026 (S.D. = .003; with a range from .020 to .030). The difference in mean pseudo-$R^2$ values indicates that the positivist model
reduces a greater proportion of the error variance than the realist model and provides a better fit to each of the 20 datasets.

**Favorable Decisions by Religious Group Status: Comparing Measures**

Models 1 and 3 in Table 6 also test Hypothesis Two, which states that claimants who belong to religious minorities will be less likely to prevail in their claims than members of more mainstream faiths. The models in Table 6 conceive of religious minority status as a function of each group’s numerical size in the state where the claim was heard. As described in Chapter 3, religious groups were classified by their family membership using the Pew Forum for Religion & Public Life’s (2008) family classification for most groups and Iannaccone’s (1994) strictness typology for members of denominational Protestant faiths (not including Black Protestants). For instance, Southern Baptists were classified as Conservative Protestants, and the logit-transformed proportion of each state’s population that is comprised of Conservative Protestants was used as the measure of religious group size in Table 6.

Model 1 tests my hypothesis about religious group size with the other legal positivist hypotheses, while Model 3 controls for all other factors except for the interaction between religious group size and judicial retention system. Neither Model 1 nor Model 3 provide support for the hypothesis that religious group size has a significant impact on the likelihood that claimants will receive favorable decisions. Members of smaller faiths did not appear to be any more or less likely than members of larger faiths to earn favorable outcomes in pursuing their claims in the state appellate courts. This finding is surprising given that prior research has noted differences in outcomes based on religious affiliation (Way & Burt 1983; Brent 1999; Wybraniec

---

56 Model 3 is interpreted here instead of Model 4 because the interaction terms in Model 4 render the main effect of religious group size, which is marginally significant ($p < .10$), not interpretable independent of the interaction terms. Model 3 allows for interpretation of this main effect independent of the group size/retention system interactions.
& Finke 2001; Sisk et al. 2004). To confirm this finding, I re-specified my models using two more traditional conceptions of religious affiliation that do not take the size of each group into account. These alternate conceptions measure religious affiliation in a series of dichotomous variables, and the models employing these dichotomous variables as predictors compare religious affiliations against a single religious reference category.

The first re-specification also relies on a combination of the Pew Forum’s (2008) family classification and Iannaccone’s (1994) strictness typology, and categorizes claimants into one of five religious groups. The group that I use as the reference category in the following models is made up of Christians, not including Roman Catholics or members of Protestant sects under Iannaccone’s classification. These two latter groups each constitute additional categories. Members of new religious movements (including groups that the Pew Forum describes as “new age” faiths) and Native American faiths form the fourth category, while members of other faiths not elsewhere classified (e.g., Jews, Muslims, Hindus and Buddhists) are placed in the fifth and final category.

Table 7 on the following page tests the effects of this specification of claimants’ religious affiliation on the likelihood of favorable outcomes for claimants. Model 1 estimates these effects exclusively, while Model 2 provides controls for the form of appellate judicial retention system used where each case was heard as well as the interactions between each of the religious affiliation categories and the retention systems. As shown in Model 1, this alternate specification of religious affiliation makes little difference in predicting outcomes, as none of the effects of group membership are significant predictors of receiving favorable decisions. For the most part, the directionality of the coefficients is as expected. Odds ratios less than 1 indicate negative relationships, meaning that members of arguably the most culturally mainstream religions in
Table 7: Modeling the Likelihood of Successful Free Exercise Claims using Nominal Religious Classifications, 1997-2011

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Religious Affiliation of Claimant</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protestant Sect</td>
<td>.913</td>
<td>1.387</td>
</tr>
<tr>
<td>Catholic</td>
<td>.454</td>
<td>.081</td>
</tr>
<tr>
<td>New Religious Movement/</td>
<td>.527</td>
<td>.079</td>
</tr>
<tr>
<td>Native American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Faith</td>
<td>1.095</td>
<td>2.537</td>
</tr>
<tr>
<td><strong>Judicial Retention System</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>.499</td>
<td></td>
</tr>
<tr>
<td>Competitive Election</td>
<td>1.792</td>
<td></td>
</tr>
<tr>
<td><strong>Interaction: Judicial Retention x Religious Affil.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election x Protestant Sect</td>
<td>3.159</td>
<td></td>
</tr>
<tr>
<td>Retention Election x Catholic</td>
<td>16.147</td>
<td></td>
</tr>
<tr>
<td>Retention Election x NRM/Native American</td>
<td>1.446</td>
<td></td>
</tr>
<tr>
<td>Retention Election x Other Faith</td>
<td>.035</td>
<td></td>
</tr>
<tr>
<td>Competitive Election x Protestant Sect</td>
<td>.170</td>
<td></td>
</tr>
<tr>
<td>Competitive Election x Catholic</td>
<td>9.254</td>
<td></td>
</tr>
<tr>
<td>Competitive Election x NRM/Native American</td>
<td>12.800</td>
<td></td>
</tr>
<tr>
<td>Competitive Election x Other Faith</td>
<td>.135</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>.155***</td>
<td>.139*</td>
</tr>
<tr>
<td>McFadden Pseudo-$R^2$ (mean of 20 imputations)</td>
<td>.015</td>
<td>.118</td>
</tr>
</tbody>
</table>

*** p<.001; ** p<.01; * p<.05; † p<.10. N = 225 using 20 imputations. Binary logistic regression coefficients expressed as logged odds.

Reference Categories:  
**Religious Affiliation of Claimant:** Christian (neither Catholic nor Protestant Sect).  
**Judicial Retention System:** No Direct Election.  
**Judicial Retention x Religious Affiliation:** No Direct Election x Christian (neither Catholic nor Protestant Sect).
America (mainline and Conservative Protestants) are more likely to receive favorable outcomes to their free exercise claims than are members of new religious movements, Protestant sects and Catholics. Members of other faiths are actually slightly more likely to receive favorable decisions than are members of the reference category. However, none of these effects are statistically significant with greater than 90 percent confidence, meaning that there is no observed effect on the outcomes of free exercise cases based on religious affiliation.

To investigate this issue further, I created a second, broader specification of religious affiliation and estimated a series of models to measure its effect on the likelihood of receiving a favorable decision. Table 8 on the next page measures religious affiliation dichotomously: claimants were classified according to whether or not they belonged to a Christian tradition. This same variable was originally used as an auxiliary variable for imputing religious group size where it was missing in Tables 6 and 7, as the text of several decisions identifies claimants as “Christian” without further specifying their religious identities.

The principal lesson of Table 8 is that the religious affiliation is still not a significant predictor of case outcomes when comparing Christians and non-Christians. When controlling for the type of legal test used and the substantive basis of the case in Model 1, claims raised by Christians were not significantly more or less likely to be upheld than were claims raised by non-Christians. This remains true in the fully-specified Model 2 as well. In Model 3, there are no significant interactions between the dichotomous religious affiliation of the claimant (Christian v. non-Christian) and the judicial retention system operating in the state where the case is heard. Based on the nature of the significant interaction described earlier, it appears that religious group size, not religious identity alone, is responsible for the significant interaction observed on the retention-election interaction term in Table 6.
Table 8: Modeling the Likelihood of Successful Free Exercise Claims using Dichotomous Religious Classification, 1997-2011

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of Balancing Test</td>
<td>3.911**</td>
<td>3.755*</td>
<td>3.689*</td>
</tr>
<tr>
<td>Christian Claimant</td>
<td>.874</td>
<td>.879</td>
<td>.538</td>
</tr>
<tr>
<td>Case Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/Child Welfare Disputes</td>
<td>.454</td>
<td>.484</td>
<td>.494</td>
</tr>
<tr>
<td>Disputes over Positive Gov’t. Policies</td>
<td>.419</td>
<td>.402</td>
<td>.421</td>
</tr>
<tr>
<td>Other Disputes</td>
<td>.392†</td>
<td>.407</td>
<td>.427</td>
</tr>
<tr>
<td>Metro Area Minority Pop Size (Log)</td>
<td>1.105</td>
<td>1.093</td>
<td></td>
</tr>
<tr>
<td>Proportion of GOP Judges</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.121</td>
<td>2.270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judicial Retention System</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>.753</td>
<td>.366</td>
<td></td>
</tr>
<tr>
<td>Competitive Election</td>
<td>1.765</td>
<td>1.500</td>
<td></td>
</tr>
<tr>
<td>Statewide Importance of Religion</td>
<td>.125</td>
<td>.110</td>
<td></td>
</tr>
<tr>
<td>Interaction: Judicial Retention x Religious Affil.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election x Christian Claimant</td>
<td>3.296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competitive Election x Christian Claimant</td>
<td>1.508</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>.093***</td>
<td>.060*</td>
<td>.085</td>
</tr>
<tr>
<td>McFadden Pseudo-$R^2$ (mean of 20 imputations)</td>
<td>.076</td>
<td>.107</td>
<td>.117</td>
</tr>
</tbody>
</table>

*** $p<.001$; ** $p<.01$; * $p<.05$; † $p<.10$.  
N = 225 using 20 imputations. Binary logistic regression coefficients expressed as odds ratios.

Reference Categories:  **Case Type:** Prisoners’ Rights/Land Use Disputes.  **Judicial Retention System:** No Direct Election.  **Judicial Retention x Religious Affiliation:** No Direct Election x Claimant is Not Christian.
The results reported in this section contrast with the finding from past research that mainline Protestants tend to fare better in their claims than members of other faiths (Brent 1999; Wybraniec & Finke 2001; Sisk et al. 2004). What might account for this discrepancy? I believe that the reason religious affiliation does not predict case outcomes in my models is the same reason I observed a lower success rate than was found in other social scientific studies of free exercise cases: I employed a more narrower definition of what constitutes a free exercise claim when locating cases. As discussed in Chapter 3, I did not include cases involving disputes that religious groups enjoy under the law but which other groups in society may not enjoy (e.g., tax exemptions and exemption from anti-discrimination laws). However, I coded cases of the latter type in the course of my content analysis and am able to test whether the inclusion of these cases alters the relationship between religious affiliation and claimant success. Analysis of all 453 cases coded initially (available upon request) indicates that under this broader conception of the meaning of religious free exercise, larger religious groups are significantly more likely to prevail in their claims than are smaller groups on a bivariate basis (odds ratio = 1.183, p < .05). However, this association becomes non-significant when accounting for claims’ substantive bases. This finding indicates that the religious affiliation effect observed in this broader model is spurious, and leads me to conclude that the religious affiliation effects reported in previous studies may be due in part to a broader conception of what constitutes a free exercise claim.

**Favorable Decisions by Case Type**

In the last set of models in this chapter, I test whether free exercise claims involving prisoners’ rights and land use disputes are more likely to be resolved in favor of the claimant than are other types of claims (Hypothesis Three-B). I test this hypothesis using data from cases
decided after September 2000, since I hypothesize that effects will only be presented after that time. As mandated by Congress following the passage of the Religious Land Use and Institutionalized Persons Act (RLUIPA) in 2000, cases involving these issues that are heard at the state level should be evaluated using the compelling interest test, a form of balancing test. The analysis in this chapter supports the hypothesis that claimants are more likely to prevail when the judges adjudicate their claims using a balancing test than when using a rational-basis test. Since cases involving prisoners’ rights and land use disputes that both involve free exercise claims are to be evaluated using a type of balancing test under RLUIPA, it follows that the likelihood of claimant successes in these types of cases should be greater than in other cases.

I test my hypothesis in the models in Table 9 (shown on the following page) first by estimating the likelihood of success in different types of cases compared to successes in prisoners’ rights and land use cases (Model 1). I then control for the form of legal test used to evaluate different claims in Model 2. Finally, I control for all other predictors in Models 3 and 4 (excluding the interactions between religious group membership size and judicial retention system in Model 3, and including these interactions in Model 4). These models provide limited support for Hypothesis Three-B. Religious free exercise claims involving prisoners’ rights and disputes over land use appear more likely to be upheld than cases of any other kind. Differences between cases of these types and two additional categories of case types (education and child welfare disputes on the one hand, and disputes involving positive government requirements, e.g., requirements to obtain immunizations and government-issued licenses on the other) were not statistically significant. However, the model indicates that cases involving matters other than these – a combination of cases populated largely by disputes over criminal convictions,
Table 9: Modeling the Likelihood of Successful Free Exercise Claims, 2000-2011

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use of Balancing Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Balancing Test</td>
<td>3.443*</td>
<td>3.890*</td>
<td>4.447*</td>
<td></td>
</tr>
<tr>
<td><strong>Religious Group Size</strong> (Proportion of State Population, Logit)</td>
<td>.916</td>
<td>.696</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Case Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/Child Welfare Disputes</td>
<td>.336</td>
<td>.344</td>
<td>.368</td>
<td>.351</td>
</tr>
<tr>
<td>Disputes over Positive Gov’t. Policies</td>
<td>.336</td>
<td>.292</td>
<td>.264</td>
<td>.304</td>
</tr>
<tr>
<td>Other Disputes</td>
<td>.355†</td>
<td>.381†</td>
<td>.368</td>
<td>.430</td>
</tr>
<tr>
<td><strong>Metro Area Minority Pop Size (Log)</strong></td>
<td>1.069</td>
<td>1.058</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Proportion of GOP Judges</strong></td>
<td>2.849</td>
<td>2.795</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Judicial Retention System</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>.596</td>
<td>2.075</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competitive Election</td>
<td>.979</td>
<td>3.628</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Statewide Importance of Religion</strong></td>
<td>.527</td>
<td>.717</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Interaction: Judicial Selection x Religious Group Size</strong></td>
<td>1.649</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election x Religious Group Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competitive Election x Religious Group Size</td>
<td>1.696</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>.270***</td>
<td>.111***</td>
<td>.039</td>
<td>.015†</td>
</tr>
<tr>
<td>McFadden Pseudo-$R^2$ (mean of 20 imputations)</td>
<td>.040</td>
<td>.080</td>
<td>.116</td>
<td>.154</td>
</tr>
</tbody>
</table>

*** $p<.001$; ** $p<.01$; * $p<.05$; † $p<.10$. N = 178 using 20 imputations. Binary logistic regression coefficients expressed as odds ratios.

Reference Categories: **Case Type:** Prisoners’ Rights/Land Use Disputes.  
**Judicial Selection System:** No Direct Election.  
**Judicial Selection x Religious Group Size:** No Direct Election x Religious Group Size
particularly the use of narcotics, as well as disputes involving the constitutionality of laws and other issues – are significantly less likely to be upheld than are cases in the reference category (on a marginal basis; \( p < .10 \)).

Even more striking than this bivariate association is that it remains significant after controlling for judges’ use of a balancing test to evaluate claims in Model 2. The odds that cases requiring the use of a balancing test under RLUIPA will be upheld are nearly 62 percent higher than for cases in the omnibus “other case type” category, after controlling for the use of the legal test mandated by RLUIPA (\( p < .10 \)).\(^57\) The statistical significance of this association is attenuated in subsequent models, however. While this association remains substantively large in the fully-specified Model 4 (the odds that prisoners’ rights and land use free exercise cases will be resolved in favor of the claimant are still 57 percent higher than for “other cases,” all else being equal), the association falls short of statistical significance at the 90 percent confidence level.

The finding in Model 2 is surprising given that I had expected that controlling for the use of the balancing test in prisoners’ rights and land use disputes, mandated under RLUIPA for evaluating these types of cases, would attenuate the relationship between case outcomes and claims made of this type. These findings suggest that prisoners’ rights and/or land use cases might be decided in favor of claimants with greater frequency for reasons other than the legal test used to adjudicate these claims. For instance, even though Brent (1999) argued that prisoners should be less likely to prevail in their claims, Lupu (1998:591) argued that the types of claims they raise tend to involve the most basic denials of religious freedom, and may thus be more likely to be upheld. Likewise, perhaps claims over religious groups’ ability to use their land as they see fit are more conducive to success than, say, claims over the constitutionality of laws on

\(^{57}\) This calculation is as follows: \((.381 - 1) \times 100 = -61.9\).
free exercise grounds. While the association between case type and case outcomes does not remain statistically significant in the fully-specified Model 4, the fact that it is marginally significant after accounting for the use of a balancing test suggests that successes in prisoners’ rights and land use claims is not merely a function of how judges evaluate claims of these kinds.

Summary

The analysis in this chapter addressed the question of what predicts outcomes favorable to claimants in state-level religious free exercise decisions. Binary logistic regression models estimated in this chapter tested six hypotheses about factors I expected would influence the outcomes of these cases. Three of the hypotheses concerned potential predictors of successes based on neutral application of the law (legal positivist hypotheses), while the other three concerned the characteristics of judges and the structural constraints in which they operate (legal realist hypotheses). In short, I do not find that many of the variables hypothesized to influence case outcomes were significantly related to these outcomes. This is not surprising given my descriptive findings in Chapter 4, which found that claimant success in free exercise cases was very rare; only 12 percent of claims were upheld in the state appellate courts between mid-1997 and 2011.

Nevertheless, I did find consistent, powerful support for one key hypothesis: when judges apply balancing tests to the facts of free exercise cases, they are significantly more likely to rule in favor of the claimants than when they apply rational-basis tests. The choice to evaluate a claim based on a legal test that balances (a) the severity of government-imposed burdens on religion against (b) the gravity of the government’s interest in imposing those burdens is consistently associated with a higher likelihood of claimant success compared to the use of a
rational-basis test. It may appear self-evident that cases decided using balancing tests should have greater success rates than those decided under the alternate test. Yet the low overall success rates for claims heard under both types of tests as reported in Chapter 4 suggests that even when a balancing test is used, be it the compelling interest test articulated in the Supreme Court’s 1963 *Sherbert v. Verner* opinion or another form of state-level balancing test, the odds are still stacked heavily against the claimants. In short, even a finding of this type is noteworthy when the odds of success for claimants are so slight.

The models revealed that in general, few of these hypotheses were supported when holding other factors constant. For instance, the substantive basis of the claim influenced the likelihood of outcomes favorable to the claimants in some instances, but not when controlling for other factors. These models lent somewhat greater support to the hypothesis that the level of public scrutiny judges face (as measured by whether public affirmation is required in order for judges to retain their seats at the end of their term on the bench) may play a role in how they rule, depending on the representation of the claimant’s religious group in the population. The robustness of this finding, however, is somewhat questionable, as it was not replicated using the smaller pool of cases examined in Table 9.

The findings in this chapter prompt an additional question: why is the association between case type and likelihood of success (at least between one class of cases and the reference group) marginally significant in Model 2 of Table 9, even after controlling for the use of a balancing test? This is an intriguing puzzle given that the logic of Hypothesis Three-B was that the use of a balancing test, which should be more likely in prisoners’ rights and land use cases than in others due to federal legislation, would be responsible for higher likelihood of success in these cases. Yet Model 2 demonstrates that this is not the case. While the “other”
case type category is too heterogeneous to draw conclusions about it, the nature of some of the
cases in the reference category may provide some insight as to why this association holds in
Model 2. Cases involving prisoners’ rights tend to involve issues such as access to religious
rituals and materials and participation in other religious practices while incarcerated.\(^{58}\) It may be
that denying claimants these things is seen by courts as more burdensome to claimants’ religious
practice and is more easily remedied than, say, holding a law unconstitutional on free exercise
grounds or allowing someone to drive without a license.

This chapter explored the factors that influence whose free exercise claims prevail in the
state courts and whose claims do not. It also highlighted the critical importance of the legal tests
judges employ in deciding who prevails. The next chapter turns our attention toward the factors
that promote the use of balancing tests in evaluating free exercise claims.

\(^{58}\) See, for example, *Hyde v. Fisher*, 203 P.3d 712 (Idaho 2009); *Meggett v. Pennsylvania Dept. of Corrections* 892
A.2d 872 (Pennsylvania 2006).
CHAPTER 6: MULTIVARIATE ANALYSIS OF LEGAL TEST USE

Introduction

This chapter explores my second basic research question, one which has attracted far less attention in past research on religious free exercise disputes: what are the factors that predict the type of legal test judges use in evaluating free exercise claims? The previous chapter highlighted that the single most consistent predictor of outcomes in state appellate free exercise cases is the type of legal test on which judges base their rulings. The use of a balancing test (which accounts for the severity of burdens on claimants’ free exercise rights) leads to a greater likelihood of claimant success than the use of rational-basis tests (which do not account for these burdens).

This finding should not be a surprise, as legal scholars have long argued that the use of different tests reflects two very different definitions of the constitutional scope of the free exercise of religion in America (McConnell 1990a; Tushnet 1993; Laycock 1994). Jelen (2010) defines these two contrasting understandings of free exercise rights in terms of a conflict between communalism and libertarianism. Communalist advocates of the rational-basis test argue that free exercise rights may be limited in the face of majority preferences. Justice Scalia’s majority opinion in Employment Division v. Smith (1990), which required the use of a rational-basis test at all levels of the judiciary where federal constitutional free exercise protections were at issue, is an endorsement of this view. In contrast, libertarian advocates argue that free exercise protections should be interpreted to generally exempt religious action from government regulation unless it is absolutely necessary to do so. This view is reflected in the logic of the compelling interest test articulated in the Court’s Sherbert v. Verner (1963) decision, the predominant form of balancing test used in state courts: if claimants show that their religious
exercise has been substantially burdened, the courts must balance the weight of this burden against the government’s interest in imposing it.

Neither form of legal test dominates state free exercise jurisprudence today. As shown in Chapter 4, 56 percent of state appellate free exercise cases were decided using some form of balancing test, while the remaining 44 percent were decided using a rational-basis test. Given the significant difference in the outcomes of cases depending on the type of test judges employ, it is worthwhile to consider the factors that promote the use of one form of test or the other. As far as I am aware, no previous study has attempted to predict whether judges will use a balancing test or a rational-basis test to evaluate a free exercise claim. The results of this chapter make a significant contribution in this regard; understanding the forces that predict legal test use in these cases will give us a more complete sense of the process by which judges reach decisions in free exercise cases.

This chapter is structured in two parts. First, I describe results from a series of binary logistic regression models that treat the choice of legal test used as a dichotomous dependent variable. A value of zero on this variable indicates that judges used a rational-basis test to evaluate a given claim, while a value of one indicates the use of some form of balancing test. This section describes whether hypothesized effects of predictors of balancing test use are supported by multivariate analysis, as well as whether other factors help predict the type of test judges use. The second section presents analysis that addresses both of the dissertation’s overarching research questions at once: what predicts judicial use of balancing tests in free exercise cases, and what predicts whether claimants will receive favorable or unfavorable decisions from the courts? The analysis in this latter section tests all of my hypotheses – those about test use as well as about case outcomes – in a structural equation modeling framework.
This approach permits me to explore the simultaneous effects of an array of predictors on the probability that judges will use a balancing test to evaluate claims as well as the probability that cases will be resolved in favor of free exercise claimants.

I test five hypotheses about factors predicting the use of a balancing test to evaluate free exercise claims. As in Chapter 5, these hypotheses are a combination of those suggested by legal positivist and legal realist logics. The two positivist hypotheses I test in this chapter are:

- After the passage of RLUIPA in 2000, claims involving prisoners’ religious rights or religious land use disputes should be more likely to be adjudicated using a balancing test than claims over other substantive issues (Hypothesis Three-A); and
- Judges will be more likely to use balancing tests to adjudicate claims in states where state supreme courts and legislative actions mandate its use (Hypothesis Four).

These two positivist hypotheses are based on the proposition that variations in outcomes are due to logical, neutral application of legal rules. They posit differences in legal test use based upon the force of precedent and legislative mandate. Hypotheses based on the logic of legal realism, on the other hand, suggest that judges’ attitudes and the structural constraints in which they operate influence their behavior. In this case, these factors will influence the type of legal test judges use in evaluating free exercise claims. The realist hypotheses effectively apply the lessons learned about the effects of attitudes and judicial maximizing behavior on case outcomes, as described in earlier chapters, to understanding how judges select the legal test they will use to evaluate a free exercise claim. The realist hypotheses I test in this chapter are as follows:
- Use of a balancing test should be less likely as the Republican composition of the panel of judges who heard a free exercise claim increases (Hypothesis Five-A);

- In states where judges do not require public approval to retain their seats, claims made by members of minority faiths should be more likely to be adjudicated using a balancing test than in states where judges face popular referenda (Hypothesis Six-A); and

- Free exercise claims will be more likely to be evaluated using a balancing test than a rational-basis test as the population of religious minorities in the locality where the claims are heard increases (Hypothesis Seven-A).

**Predicting Use of Balancing Tests to Evaluate Free Exercise Claims**

In order to test these hypotheses, I estimate a series of logistic regression models predicting the likelihood that judges will use balancing tests to evaluate free exercise claims. The models in Table 10 on the following page present estimates from three logistic regression models predicting balancing test use based on a number of hypothesized covariates, as well as an indicator of state-level religious salience. Model 1 explores the effect of past state precedent on the use of this type of test independently, a test of Hypothesis Four. Model 2 controls for all hypothesized predictors and a number of additional controls, though the interactions between religious group size and state judicial retention system type are omitted so that the main effects of these two variables may be interpreted. Model 3 replicates Model 2 and adds these interaction terms.

The results of Model 1 support my hypothesis that the existence of judicial precedents and legislative mandates for the use of a balancing test was significantly and positively associated with its use in adjudicating free exercise cases ($p < .05$). The association between
Table 10: Modeling the Likelihood of Balancing Test Use, 1997-2011

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precedent for Balancing Test Use</td>
<td>1.882*</td>
<td>1.834†</td>
<td>1.821†</td>
</tr>
<tr>
<td><strong>Religious Group Size</strong> (Proportion of State Population, Logit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/Child Welfare Disputes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disputes over Positive Gov’t. Policies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Disputes</td>
<td>.936</td>
<td>.930</td>
<td></td>
</tr>
<tr>
<td>Metro Area Minority Pop Size (Log)</td>
<td>1.100</td>
<td>1.106</td>
<td></td>
</tr>
<tr>
<td>Proportion of GOP Judges</td>
<td>.429†</td>
<td>.408†</td>
<td></td>
</tr>
<tr>
<td><strong>Judicial Retention System</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>2.479*</td>
<td>2.241</td>
<td></td>
</tr>
<tr>
<td>Competitive Election</td>
<td>3.920**</td>
<td>4.071*</td>
<td></td>
</tr>
<tr>
<td>Statewide Importance of Religion</td>
<td>.002***</td>
<td>.002***</td>
<td></td>
</tr>
<tr>
<td><strong>Interaction: Judicial Selection x Religious Group Size</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election x Religious Group Size</td>
<td></td>
<td></td>
<td>.940</td>
</tr>
<tr>
<td>Competitive Election x Religious Group Size</td>
<td></td>
<td></td>
<td>1.010</td>
</tr>
<tr>
<td>Constant</td>
<td>-.134</td>
<td>7.226</td>
<td>7.564</td>
</tr>
<tr>
<td>McFadden Pseudo-$R^2$ (mean of 20 imputations)</td>
<td>.017</td>
<td>.086</td>
<td>.091</td>
</tr>
</tbody>
</table>

*** $p$.001; ** $p$.01; * $p$.05; † $p$.10. N = 225 using 20 imputations. Binary logistic regression coefficients expressed as logged odds.

Reference Categories: **Case Type:** Prisoners’ Rights/Land Use Disputes.
**Judicial Retention System:** No Direct Election.
**Judicial Retention x Religious Group Size:** No Direct Election x Religious Group Size.

these variables is attenuated somewhat in Models 2 and 3, in which precedent is a marginally significant positive predictor of balancing test use. There are a number of possible reasons why precedent alone is not the strongest predictor of balancing test use when holding other factors constant. One reason may be that the dataset includes a number of cases that were precedent-
setting in and of themselves; in these cases, state supreme court justices established a precedent that one or the other type of test should be used when interpreting the state’s constitutional free exercise protections. These are instances of judicial decision-making that could not be based on existing state precedent, since the matter of how to evaluate free exercise disputes under these states’ constitutions had not yet been resolved. Another reason may be that the dataset contains cases from intermediate appellate courts, which, in the absence of guidance from their states’ supreme courts, decided to employ one form of test or the other independently. Nearly half of all states fall into the category of states that (a) had not enacted legislation requiring the use of a balancing test in adjudicating free exercise disputes and (b) had no judicial precedent for evaluating free exercise disputes between 1990 (when the Supreme Court established the rational-basis test for interpretations of federal constitutional claims in state courts in its \textit{Employment Division v. Smith} decision) and 2010 (Durham and Smith 2010:93). The absence of precedent on how to interpret state constitutional provisions in these states created the opportunity for lower appellate courts to apply whichever test they deemed appropriate.

I test additional hypotheses about the predictors of balancing test use in Models 2 and 3 in Table 10. Results from the tests of three hypotheses (Hypotheses Five-A, Six-A and Seven-A, all hypotheses emerging from legal realist logics) can be summarized here. Hypothesis Five-A stated that balancing test use is more likely in cases decided by smaller proportions of judges affiliated with the Republican Party. Both Models 2 and 3 provide support for this hypothesis. In the fully-specified Model 3, the predicted odds that a balancing test will be used to evaluate a free exercise claim are about 59 percent lower when the claim is heard by a panel of entirely

Republican-affiliated judges than when the claim is heard by entirely non-Republican judges.\textsuperscript{60} In short, the less Republican the composition of the bench hearing a claim is, the more likely the claim will be evaluated using a balancing test rather than a rational-basis test. This is intriguing given that the bivariate analysis in Chapter 4 did not detect this association, which appears to have been suppressed at the bivariate level. Republican judges appear to follow Justice Scalia in taking a more communalist view of free exercise rights than do their Democratic counterparts, using Jelen’s (2010) distinction in interpreters’ orientations toward these rights. The negative association between the proportion of Republicans on the bench and the likelihood of balancing test use is marginally significant with greater than 90 percent confidence ($p < .10$).

I test whether different judicial retention systems influence judges to select different legal tests (given claimants’ minority status) in Model 3. This hypothesis is tested using interaction terms, which posit effects on balancing test likelihood as a multiplicative function of (a) the form of judicial retention system at the state level and (b) the size of the religious group to which the claimant belongs in that state (expressed as a logit-transformed proportion of the state’s population). My analysis finds no support this hypothesis; judges do not choose different legal tests based on the interaction between claimants’ religious affiliations and the judicial retention systems in each state.

However, Model 2 reveals that judicial retention systems themselves do influence the form of tests judges use, controlling for all other factors (including past precedent). Judges are significantly more likely to employ balancing tests in evaluating free exercise claims when they retain their seats on the bench through public referenda. As shown in Model 2, the odds that a claim will be adjudicated under a balancing test are nearly 150 percent higher when it is made in a state where judges retain their seats through non-competitive retention elections than when

\textsuperscript{60} This calculation was performed as follows: $(.408 - 1) \times 100 = -59.2$. 

123
The use of balancing tests is even more likely in states where judges run against other candidates for their seats in competitive elections. Model 2 shows that the odds that a case will be decided using a balancing test are nearly 300 greater in when heard in states like this than when judges are not retained through direct election ($p < .01$).

Finally, I test the hypothesis that that use of a balancing test would be more likely in metropolitan or micropolitan areas populated by greater numbers of religious minorities in Models 2 and 3. As measured by the natural log of the local population estimate that is not Protestant, Catholic nor Mormon, minority population size exerts no significant impact on the likelihood that judges in an area will elect to use this test in evaluating free exercise claims. It would appear that greater exposure to religious minorities in judges’ lives does not imbue them with a more libertarian orientation to the nature of free exercise rights. Limiting the number of observations to only those cases resolved after September 2000, as the models in Table 11 do (see page 125), renders the association marginally significant ($p < .10$) and in the hypothesized direction. However, given the longer temporal window and greater number of cases in the Table 10 analysis, I am inclined to put more stock in the hypothesis test in Table 10 that fails to support this hypothesis.

Before turning to the last hypothesis test of this section, it is worth noting that Table 10 reports an additional finding that I had not hypothesized. I included a control for statewide religious salience using a measure that indicates the proportion of states’ populations that view religion as “very important” in their lives (Pew Forum 2008). This measure of religious salience was not found to be a significant predictor of free exercise case outcomes in Chapter 5. Yet

\[ (2.479 - 1) \times 100 = 147.9 \text{ percent.} \]

\[ (3.920 - 1) \times 100 = 292.0 \text{ percent.} \]
remarkably, this measure is not only a significant predictor of the type of test used, but it is the only predictor in Table 10 that is found to be associated with legal test use with greater than 99.9 percent confidence. “Importance of religion” among the population of states where claims are heard is a strong, negative predictor of balancing test use. Judges are significantly more likely to use rational-basis tests to evaluate free exercise claims as the proportion of the population in their state that views religion as “very important” in their lives increases, net of other factors controlled for in Models 2 and 3. This is particularly unexpected given the directionality of the relationship, as one might expect to see broader use of tests that are more protective of religious freedom in states where religion is a more substantial cultural force in the population. I will discuss the implications of this finding in greater depth in the following chapter.63

Finally, in order to test whether different types of cases are significantly more likely to be evaluated under a balancing test, I estimated two additional logistic regression models, the results of which are reported in Table 11 on the next page. This final hypothesis (Hypothesis Three-A) states that cases involving claims protected under the Religious Land Use and Institutionalized Persons Act (RLUIPA) should be more likely to be evaluated under a balancing test than other types of cases. Because RLUIPA came into force in September 22, 2000, only cases decided after that date were included in the analysis in this Table. Model 1 compares the likelihoods of balancing test use among different types of claims.

63 As shown in Appendix F, the statewide importance of religion measure is correlated with the competitive elections judicial retention system indicator at $\rho = .470$. I estimated two separate models removing statewide importance of religion and the judicial retention system indicators respectively (available upon request); the models were otherwise identical to those shown in Model 2 of Table 10. Statewide importance of religion remained a significant negative predictor of balancing test use in the model excluding judicial retention systems, but neither judicial retention system indicator were significant predictors in the converse model. This may suggest the presence of a suppressor effect, but given Gordon’s (1968) warning about the partialling fallacy, committed when including correlated variables in models predicting some outcome, we should be cautious in drawing too strong a conclusion about the effect of judicial retention systems on balancing test use.
### Table 11: Modeling the Likelihood of Balancing Test Use, 2000-2011

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case Type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/Child Welfare Disputes</td>
<td>.790</td>
<td>.835</td>
</tr>
<tr>
<td>Disputes over Positive Gov’t. Policies</td>
<td>1.692</td>
<td>1.729</td>
</tr>
<tr>
<td>Other Disputes</td>
<td>.673</td>
<td>.626</td>
</tr>
<tr>
<td><strong>Precedent for Balancing Test Use</strong></td>
<td>.852</td>
<td></td>
</tr>
<tr>
<td><strong>Religious Group Size</strong> (Proportion of State Population, Logit)</td>
<td>1.087</td>
<td></td>
</tr>
<tr>
<td><strong>Metro Area Minority Pop Size (Log)</strong></td>
<td>1.177†</td>
<td></td>
</tr>
<tr>
<td><strong>Proportion of GOP Judges</strong></td>
<td>.342†</td>
<td></td>
</tr>
<tr>
<td><strong>Judicial Retention System</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>2.328</td>
<td></td>
</tr>
<tr>
<td>Competitive Election</td>
<td>4.058†</td>
<td></td>
</tr>
<tr>
<td><strong>Statewide Importance of Religion</strong></td>
<td>.001***</td>
<td></td>
</tr>
<tr>
<td><strong>Interaction: Judicial Selection x Religious Group Size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election x Religious Group Size</td>
<td>.902</td>
<td></td>
</tr>
<tr>
<td>Competitive Election x Religious Group Size</td>
<td>.916</td>
<td></td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>1.773*</td>
<td>17.963†</td>
</tr>
<tr>
<td><strong>McFadden Pseudo-(R^2)</strong> (mean of 20 imputations)</td>
<td>.015</td>
<td>.100</td>
</tr>
</tbody>
</table>

*** \(p<.001\); ** \(p<.01\); * \(p<.05\); † \(p<.10\). N = 178 using 20 imputations. Binary logistic regression coefficients expressed as logged odds.

Reference Categories: **Case Type:** Prisoners’ Rights/Land Use Disputes.  
**Judicial Retention System:** No Direct Election.  
**Judicial Retention x Religious Group Size:** No Direct Election x Religious Group Size.

Surprisingly, I find no evidence that case types ostensibly controlled by RLUIPA are evaluated using balancing tests more frequently than other types of cases; this indicates that Hypothesis Three-A is not supported by my analysis. The fully-specified Model 2 yields the same finding; there appears to be no significant link between case type and the use of a balancing...
test. To explore this finding in more detail, I estimated an additional binary logistic regression model collapsing categories of claim types into a dichotomy: cases involving claims that should be theoretically protected under RLUIPA (i.e., prisoners’ rights and land use claims) and all other claims. Even in this simple bivariate model (available upon request), I observed no significant distinction between case types regarding use of a balancing test. The association also remains non-significant when the effect of precedent is removed from Model 2.

This finding suggests that while claims evaluated under a balancing test are more successful than claims evaluated under a rational-basis test, federal legislative efforts to expand the use of the balancing test may not have had their intended effect at the state level. While the number of observations in the analysis in Table 11 is not large (N = 178), which may increase the possibility of Type II error, it is also possible that claimants, lawyers and judges simply failed to recognize RLUIPA protections that should have applied to various claims or else chose to overlook them.

A closer reading of claims that should have been adjudicated under a balancing test due to RLUIPA suggests that this may be the case. For instance, in Austin v. Crosby (866 So.2d 742 [2004]), a case heard in Florida’s Fifth District Court of Appeal, a Wiccan inmate petitioned for a writ of mandamus to allow her to participate in and perform a number of religious rituals while incarcerated. Even though the case was decided in early 2004, several years after the passage of RLUIPA, the court explicitly relied on the rational-basis tests put forward in the U.S. Supreme Court’s 1987 Turner v. Safley and O’Lone v. Estate of Shabazz decisions on prisoners’ free exercise rights in order to make its ruling. RLUIPA was not referenced in this opinion. Likewise, in Shepherd Montessori Center Milan v. Ann Arbor Charter Township (783 N.W.2d 695 [2010]), the Michigan Supreme Court applied the rational-basis test put forward in
Employment Division v. Smith (1990) as well as Church of the Lukumi Babalu Aye v. City of Hialeah (1993) in holding that a land use ordinance was not applied in a discriminatory manner, though RLUIPA’s compelling interest test could conceivably have been used. In short, these results underscore the fact that attempts to legislate the use of some form of balancing test do not necessarily lead to its universal implementation. Legislative action mandating the use of a balancing test appears insufficient in guaranteeing that such a test will be used in the courts.

Taken together, the findings in this section on the predictors of balancing test use indicate that precedent alone does not determine the standard by which judges evaluate religious free exercise claims. While past precedents for the use of this test do increase the likelihood that it will be used, a number of factors concerning judges’ ideologies and the context in which cases are heard also predict balancing test usage. Building upon research indicating that Republican judges are less likely to rule in favor of claimants in civil liberties cases (Segal and Cover 1989; Marshall and Ignagni 1994), I find that panels of judges are less likely to employ a legal test that offers claimants greater protection as the proportions of Republicans on the panels increase. Further, I had hypothesized that greater public accountability, measured by whether the public has a say in allowing judges to retain their seats, would be associated with reduced likelihoods that judges will use balancing tests. While I did not find support for this hypothesis, I found that greater public accountability is associated with a greater likelihood of balancing test usage. This finding is counter to the logic of this hypothesis, but it suggests that judges are more likely to advocate tests that protect citizens’ religious rights when they rely on the public in keeping their seats on the bench.

64 The opinion references RLUIPA protections in a footnote, but the court does not invoke them in making its decision in this case.
Simultaneous Prediction of Balancing Test Use and Case Outcomes

The analysis reported in Chapter 5 tested hypotheses about the factors predicting decisions favorable to claimants in religious free exercise claims, while the analysis presented in the previous section did the same regarding the use of balancing tests to adjudicate these claims. I conclude my quantitative analysis of state-level free exercise data by putting together the lessons of these two chapters, by using a structural equation modeling framework to understand how well both sets of models fit the data. This section begins with the estimation of models based on equations that both predict test use and case outcomes. After comparing estimates from these models, I fit a trimmed set of models that more parsimoniously reflects the nature of the primary factors influencing these endogenous variables. I conclude by trying to determine whether this more parsimonious model better reflects the structure of the data.

I used the Mplus 6 statistical package to estimate the structural equation models in this section of the chapter. Mplus is especially well-suited to model equations with dichotomous dependent or endogenous variables, an important consideration given that both legal test selection and case outcomes (as are several of my hypothesized predictors). The means- and variance-adjusted robust weighted least squares (WLSMV) estimator in Mplus models these kinds of variables according to their conceptual continuous latent distributions (Muthén, du Toit and Spisic 1997), which allows them to be treated much like continuous variables in structural equation modeling.

Use of the WLSMV estimator instead of a maximum-likelihood (ML) estimator yields probit rather than logistic regression coefficients, which are based on the distribution of the cumulative standard normal curve. Interpretation of these coefficients in their raw form is less substantively meaningful or intuitive than interpretation of logistic regression coefficients, but
the statistical significance and directionality of the relationships they describe are both readily interpretable (Pampel 2000:54-61). Also, unlike maximum likelihood estimation (the basis for estimating logistic regression models) using categorical endogenous variables, it is currently possible using the WLSMV estimator to calculate the values of common approximate fit indices like the Steiger-Lind root mean square error of approximation (RMSEA; Steiger 2000) and the Bentler Comparative Fit Index (CFI; Bentler 1990; see also Kline 2011:204). For these reasons, I have chosen to estimate the models in this section using this estimator rather than a maximum likelihood estimator. The structural equation models estimated below are based on the same multiple imputation procedure used to estimate the logistic regression models in previous chapters. Twenty datasets imputing missing data were generated in Mplus using a Markov chain Monte Carlo algorithm, and estimates for these models were pooled from across the datasets consistent with Rubin’s (1987) combination rules.

Table 12 on the following page reports unstandardized probit regression coefficients from the structural equation model predicting differences in probabilities that (a) judges will use balancing tests to evaluate free exercise claims and (b) judges will rule in favor of the claimants. Note that these equations do not contain the interaction terms multiplying religious group size by the judicial retention system indicators. Neither of these interactions contributed to the fit of the model nor were significant predictors of change in either endogenous variable in this model; they were removed to facilitate clearer interpretation of the main effects of the religious group size and judicial retention system indicators. A path diagram portraying the structure of the model is provided in Appendix C.65

65 Note that the path diagrams in Appendix C do not contain paths recognizing correlations between exogenous variables or disturbance terms for the endogenous variables. All exogenous variables were correlated in these models but these correlation paths are omitted from the path diagrams to make visual presentation of the model
Table 12: Estimates from Structural Equation Model Predicting Balancing Test Use and Favorable Outcomes for Claimants, 1997-2011

<table>
<thead>
<tr>
<th>Variable</th>
<th>Balancing Test</th>
<th>Favorable Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precedent for Balancing Test Use</td>
<td>.433*</td>
<td>-----</td>
</tr>
<tr>
<td>Religious Group Size (Proportion of State Population, Logit)</td>
<td>.067</td>
<td>-.088</td>
</tr>
<tr>
<td><strong>Case Type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/Child Welfare Disputes</td>
<td>.250</td>
<td>-.323</td>
</tr>
<tr>
<td>Disputes over Positive Gov’t. Policies</td>
<td>.389</td>
<td>-.445</td>
</tr>
<tr>
<td>Other Disputes</td>
<td>-.055</td>
<td>-.312</td>
</tr>
<tr>
<td><strong>Metro Area Minority Pop Size (Log)</strong></td>
<td>.072</td>
<td>-.073</td>
</tr>
<tr>
<td><strong>Proportion of GOP Judges</strong></td>
<td>-.549†</td>
<td>.626</td>
</tr>
<tr>
<td><strong>Judicial Retention System</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>.555*</td>
<td>-.306</td>
</tr>
<tr>
<td>Competitive Election</td>
<td>.817**</td>
<td>-.043</td>
</tr>
<tr>
<td><strong>Statewide Importance of Religion</strong></td>
<td>-3.537**</td>
<td>-.345</td>
</tr>
<tr>
<td><strong>Use of Balancing Test</strong></td>
<td>-----</td>
<td>.485***</td>
</tr>
<tr>
<td>McKelvey and Zavoina Pseudo-$R^2$</td>
<td>.181</td>
<td>.305</td>
</tr>
</tbody>
</table>

*** $p<.001$; ** $p<.01$; * $p<.05$; † $p<.10$. N = 225 using 20 imputations. Unstandardized probit regression coefficients are reported.

Model Fit (mean values): $\chi^2 = .476$ (df = 1); RMSEA = .000; CFI = 1.000.

Reference Categories: **Case Type:** Prisoners’ Rights/Land Use Disputes.  
**Judicial Retention System:** No Direct Election.

The results of the probit regression analysis reported in Table 12 largely replicate previous findings about the nature of the relationships portrayed in the model. The existence of precedent or legislative guidelines for state-level use of the balancing test is a significant positive more straightforward. The polychoric correlation matrix describing the correlations between variables is provided in Appendix F. Probit models assume a fixed residual variance (equal to 1), so disturbance terms were also omitted.
predictor that judges will use this type of test; unlike in the single-equation models reported earlier in this chapter, the association is significant with greater than 95 percent confidence ($p < .05$). Likewise, the form of retention system judges face in keeping their seats predicts use of a balancing test; judges who must stand for noncompetitive retention elections ($p < .05$) and competitive multicandidate elections ($p < .01$) are significantly more likely to use this type of test to evaluate free exercise claims than those who do not face direct election. Negative predictors of balancing test use are, consistent with findings reported earlier in the chapter, the proportion of Republican-affiliated judges who hear a claim ($p < .10$) and the proportion of states’ populations who feel religion is “very important” in their lives ($p < .01$).

Results from the equation predicting the probability that judges will rule in favor of the free exercise claimants also mirror prior findings, though the interaction terms combining religious group size and judicial retention system (not included in the model in Table 12) failed to predict favorable decisions. Holding all else constant (except for precedent for balancing tests, which was not tested), the only predictor of favorable outcomes for free exercise claimants is the use of a balancing test rather than a rational-basis test ($p < .001$). This is consistent with my finding in Chapter 5 that balancing test use is the most robust predictor of favorable outcomes. In addition, a second model was estimated to test for case type effects using decisions made after the passage of the Religious Land Use and Institutionalized Persons Act (RLUIPA), which mandates that a balancing test be used to evaluate particular types of claims. The results of this model, which are provided in Appendix E, highlight that case types are not significant predictors of either test use or case outcomes; this is also consistent with findings from Chapters 5 and 6.
Values of a number of model fit statistics suggest that the model described in Table 12 fits the data well, though this is likely influenced by the fact that the model was estimated with just one degree of freedom. The mean model chi-square statistic for this model ($\chi^2$) across the twenty imputed datasets has a value of 10.476. This value is non-significant at one degree of freedom, which indicates that there are no significant discrepancies between covariances predicted by the model and those in the population (Kline 2011:199). The mean values of other fit statistics across the imputed datasets also indicate that the model evinces a high level of fit to the data. The value of the Comparative Index suggests this (CFI = 1.000, since the number of degrees of freedom exceeds the value of the model chi-square statistic); Hu and Bentler (1999:6) suggest that CFI values greater than .95 connote satisfactory levels of model fit. The value of the Root Mean Square Error of Approximation (RMSEA = 0.000) also suggests this, as RMSEA values less than .05 generally suggest good-fitting models (Maruyama 1998:247).

I do not test for mediation effects in this analysis, given the absence of direct effects between exogenous variables and the outcome variable (case outcomes) as found in Chapter 5 (Baron and Kenny 1986:1176). Nevertheless, it is possible for me to compare and describe the size of direct, indirect and total effects of the exogenous predictors on case outcomes using legal test use as an intervening variable. Table 13 below provides total, direct and indirect effects of these predictors (as well as the use of a balancing test) on the probability that claimants will receive favorable decisions, expressed as standardized probit coefficients. The coefficients in Table 13 demonstrate that the use of balancing test exerts the single most powerful effect on the probability that claimants will receive favorable decisions. No other predictor approaches the

---

66 These coefficients are not tested for statistical significance, as this procedure is not supported in Mplus for multiply imputed data under WLSMV estimation.
Table 13: Total, Direct and Indirect Effects of Predictors on Case Outcomes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total</th>
<th>Direct</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of Balancing Test</td>
<td>.494</td>
<td>.494</td>
<td>----</td>
</tr>
<tr>
<td>Religious Group Size</td>
<td>-.099</td>
<td>-.130</td>
<td>-.031</td>
</tr>
<tr>
<td>(Proportion of State Population, Logit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/Child Welfare Disputes</td>
<td>-.076</td>
<td>-.121</td>
<td>.045</td>
</tr>
<tr>
<td>Disputes over Positive Gov’t. Policies</td>
<td>-.088</td>
<td>-.149</td>
<td>-.061</td>
</tr>
<tr>
<td>Other Disputes</td>
<td>-.143</td>
<td>-.135</td>
<td>-.008</td>
</tr>
<tr>
<td>Metro Area Minority Pop Size (Log)</td>
<td>-.061</td>
<td>-.119</td>
<td>.058</td>
</tr>
<tr>
<td>Proportion of GOP Judges</td>
<td>.116</td>
<td>.198</td>
<td>-.082</td>
</tr>
<tr>
<td>Judicial Retention System</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>-.020</td>
<td>-.141</td>
<td>.121</td>
</tr>
<tr>
<td>Competitive Election</td>
<td>.155</td>
<td>-.020</td>
<td>.176</td>
</tr>
<tr>
<td>Statewide Importance of Religion</td>
<td>-.175</td>
<td>-.014</td>
<td>-.160</td>
</tr>
</tbody>
</table>

Note: As shown in Appendix C, all indirect effects occur through legal test use (balancing or rational-basis). Standardized probit coefficients are reported in this table.

effect that balancing test use exerts on judges’ decisions in these cases when accounting for indirect effects of exogenous predictors via balancing test use.

Given the lack of support shown for a number of hypotheses in the models described in Chapters 5 and 6, it is useful to assess whether a more parsimonious model of the predictors of balancing test use and case outcomes fits the data significantly worse than the larger model reported in Table 12. Structural equation modeling permits comparison of nested models using differences in chi-square model fit statistics. The model estimated in Table 14 on the following page estimates effects on the two endogenous variables using a smaller number of measures, which were generally found to be significant predictors of one or both of the endogenous variables in previous models. Most of the same associations previously observed are found in
Table 14: Estimates from Trimmed Structural Equation Model Predicting Balancing Test Use and Favorable Outcomes for Claimants, 1997-2011

<table>
<thead>
<tr>
<th>Variable</th>
<th>Balancing Test</th>
<th>Favorable Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precedent for Balancing Test Use</td>
<td>.434*</td>
<td>-----</td>
</tr>
<tr>
<td>Proportion of GOP Judges</td>
<td>-.435</td>
<td>-----</td>
</tr>
<tr>
<td><strong>Judicial Retention System</strong> (reference cat. = No Direct Election)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>.555*</td>
<td>-.295</td>
</tr>
<tr>
<td>Competitive Election</td>
<td>.817**</td>
<td>-.027</td>
</tr>
<tr>
<td>Statewide Importance of Religion</td>
<td>-3.589**</td>
<td>-----</td>
</tr>
<tr>
<td>Use of Balancing Test</td>
<td>-----</td>
<td>.453***</td>
</tr>
<tr>
<td>McKelvey and Zavoina Pseudo-$R^2$</td>
<td>.155</td>
<td>.246</td>
</tr>
</tbody>
</table>

*** p<.001; ** p<.01; * p<.05; † p<.10. N = 225 using 20 imputations. Unstandardized probit regression coefficients are reported. All other path coefficients for variables included in Table 12 model were constrained to 0.

Model Fit (mean values): $\chi^2 = 10.445$ (df = 13); RMSEA = .000; CFI = 1.000.

the trimmed model, though no significant association is found here between the political composition of the bench and the probability of balancing test use. According to the fit statistics associated with this more parsimonious model, the model provides a good fit with the data as well. The mean model chi-square statistic value is 10.445, which is not significantly different from a perfectly fitting model at 13 degrees of freedom. The value of the CFI for this model is greater than .95 and the value of the RMSEA is less than .05, both indicating a high degree of model fit (Hu and Bentler 1999:6). A path diagram depicting the structure of this model is provided in Appendix D; paths constrained to zero in this model were eliminated from the path diagram for easier interpretation.

Chi-square difference statistics are ordinarily calculated and tested to determine whether more parsimonious models fit the data significantly worse than more fully-specified models. A
single test for comparing chi-square fit statistics between nested models based on the WLSMV estimator is currently unavailable in Mplus for data generated by multiple imputation. However, one alternative is to calculate chi-square model fit statistics and compare them between nested models for each of the imputed datasets. In this case, I generated point estimates of each of chi-square statistics for both the full model and the more parsimonious model; I then tested the difference between them using the DIFFTEST command in Mplus (mean = 10.078 and standard deviation = 1.19, with differences of 12 degrees of freedom between each pair of models). The chi-square difference statistic was not statistically significant with greater than 90 percent confidence in any of the twenty datasets evaluated, which suggests that the trimmed model portrays the relationships between predictors and the two outcome variables just as well as the larger, less parsimonious model. In other words, it demonstrates that a number of hypothesized predictors of effects on both variables (namely, religious group size, case type and minority population size) may be disregarded without harming the ability of the model to reproduce the underlying covariance matrix.

Summary

The analysis in this chapter serves two purposes: first, to test hypotheses from legal positivist and legal realist perspectives on the predictors of judges’ usage of balancing tests to adjudicate free exercise claims, and second, to assess the fit of the models predicting both test use and case outcomes. We can draw several conclusions from the tests performed to serve these purposes. Most notably, I found support for legal realist arguments for why the balancing test might or might not be used. Judicial ideology plays a key role in determining the form of legal test judges use; Republican judges were significantly less likely to rely on balancing tests, all
else being equal. The structural constraints in which judges operate also play a role in test selection. Judges who retain their seats through some form of popular election are significantly more likely to use balancing tests in adjudicating free exercise claims than judges who are not subject to direct public scrutiny.67

These findings underscore the fact that precedent for use of the balancing test is only one of several factors that influence the type of reasoning judges employ in evaluating free exercise claims. While the models in this chapter support the proposition that judges tend to heed precedent in selecting legal tests, they also support realist arguments that judges act out of their own self-interest and according to their ideologies in deciding how to evaluate free exercise claims. In other words, while precedent matters in the judicial decision-making process, it does not determine how judges’ behavior in the state appellate courts regarding disputes over religious free exercise.

The second part of the chapter reported estimates of structural equation models predicting both balancing test use and favorable decisions for claimants in free exercise cases. This was done with an eye toward obtaining a model of predictors that best fit the data’s underlying covariance structure, as well as to minimizing exogenous variables that did not predict either balancing test use or favorable decisions. I found that a more parsimonious model removing a number of hypothesized predictors – religious group size, case type and minority population size – fits the data just as well as the model that includes these predictors. This indicates that many predictors I had hypothesized to influence case outcomes had little or no effect. However, consistent with my findings in Chapter 5, I observed that when the direct, indirect and total

---

67 However, as noted previously (see footnote 63), this effect is significant only when controlling for statewide importance of religion, which is itself substantially correlated with one of the judicial retention system indicators (competitive elections).
effects of predictor variables are considered, balancing test usage is by far the most important predictor of the direction of case outcomes.

In sum, the analysis in this chapter offers further support for my hypothesis that balancing test use is a powerful, positive predictor of favorable decisions for claimants. This chapter also contributes to emerging realist understandings of judicial decision-making, particularly from a rational-choice perspective, in that it finds several links between judicial attitudes and self-interest and how judges behave. However, I also find in my analysis that there is very little that explains variation in case outcomes other than the use of the balancing test.
CHAPTER 7: CONCLUSION

Introduction

This dissertation set out to explore the status of religious free exercise in America at the state level over the last fifteen years. Disagreement between the Supreme Court and Congress over the meaning and expansiveness of this constitutional right created two different paradigms for evaluating disputes over alleged free exercise violations in state courts. Those who make these allegations face two major uncertainties in court. First, will judges interpret constitutional free exercise protections to mean that religious and secular legal obligations ought to be balanced, such that the latter do not always take precedent over the former? Second, which side of the argument do the judges feel has greater validity in any given case? The dissertation employed quantitative analysis of court decisions in an attempt to understand why judges answer these questions in the ways they do throughout American appellate state court systems. This conclusion chapter synthesizes the results of this analysis and discusses its implications for religious free exercise in America and for further research on the subject. First, the chapter offers a recapitulation of the motivations for conducting this study and addresses why answers to its guiding research questions are important for both theoretical and practical reasons. Second, it reviews the study’s empirical findings and discusses which of the study’s arguments were supported more strongly than others by these results, and it addresses the theoretical and practical implications of these findings. Finally, it reflects on the limitations and assumptions of the study itself and offers guidance for future lines of inquiry.
Review of Study Background

Scholarly interest in the meaning and status of religious free exercise in America surged after the Supreme Court’s decision in *Employment Division v. Smith* in 1990. In this decision, the Court held that no accommodations to religious exercise in the face of a burdensome law or act of government if the law or act were neutral toward religion and applied to everyone. Legal scholars took a particular interest in the question; some argued that the Court had created a controversial new and limited definition of religious freedom (Wood 1990; McConnell 1990a; Choper 1992), while others argued that exemptions to religiously neutral laws on free exercise grounds should not be permitted and that the rational-basis test introduced in *Smith* for deciding free exercise disputes was appropriate (Marshall 1990; West 1994; Hamilton 2005). Sociologists and political scientists explored the meaning of religious free exercise differently, emphasizing patterns in the outcomes of relevant disputes in court rather than arguing that either interpretation was correct (Way and Burt 1983; Brent 1999; Wybraniec and Finke 2001; Adamczyk et al.; Sisk et al. 2004; Claborn 2011).

The Supreme Court’s 1997 decision in *City of Boerne v. Flores*, in which it effectively held that the *Smith* precedent controls how free exercise disputes based on the U.S. Constitution’s Free Exercise Clause should be resolved in state courts, marked the beginning of a period of relative stability in the status of free exercise. This state court decisions on religious free exercise analyzed in this dissertation are a product of the post-*Flores* reality. Claims in state courts based on the U.S. Constitution’s Free Exercise Clause are subject to the rational-basis test articulated in *Smith*, while these courts are free to interpret their own state constitutional religious freedom protections as they see fit. Moreover, both Congress and state legislatures have enacted laws about how all or certain types of free exercise disputes should be resolved at
this level of the judiciary. The result, as detailed in Chapter 4, is that neither the rational-basis test nor legal tests that balance religious and legal obligations dominate as the principal method judges use to evaluate free exercise disputes in the states. To my knowledge, no study has exclusively examined the body of state free exercise cases in the post-*Flores* era to assess the factors that influence the success and failure of claims.

Such an inquiry is valuable for several reasons. First, insights from religious economies theory highlight the potential ill effects on the health of the religious sphere in any society when some groups are prevented from practicing their religion as they see fit. Restrictions of this kind can limit the range of religious alternatives available to perspective adherents (Stark and Finke 2000). Though overt legal restrictions on religious groups in America are rare, incidental and inadvertent restrictions arising from religiously neutral laws can conceivably have the effect of outlawing certain religious groups. Laycock (1994) observed that had the outcomes of several key cases concerning religious groups that allegedly ran afoul of neutral and generally applicable laws been different, a number of these groups, including ISKCON (also known as the Hare Krishna movement), would have been prevented from practicing or maintaining an organizational presence. Moreover, previous studies suggest that members of religious minority groups were not only more likely to claim free exercise violations, but were less likely to prevail in their claims than members of more mainstream faiths in past decades (Wybraniec and Finke 2001; Adamczyk et al. 2004). Contemporary analysis of free exercise case outcomes and the legal rationales for those outcomes can show us the scope of these inadvertent restrictions in the states in the post-*Flores* era as well as whether they affect religious minorities disproportionately.
The inquiry is also valuable as an opportunity to test the usefulness of different theoretical perspectives from the sociology of law. Legal positivist theory views law as data to be interpreted as written, both by its interpreters in legal institutions and by scholars who examine its effects. A law’s effects on society are consequences of its content. Legal realism, on the other hand, emphasizes more strongly the human agency inherent in the interpretation and application of the law. Realism posits that interpretations of the law vary by the attitudes of the interpreter, by structural constraints placed upon the interpreter and by strategic considerations, among other reasons. Scholarship on judicial behavioralism emphasizes factors of this kind that account for variation in how judges rule across cases. This study tests a range of hypotheses based upon both perspectives, and the results of these tests contribute to a greater understanding of each perspective’s validity in explaining legal outcomes.

Findings from Analysis of Court Decisions and Their Implications

Chapters 4 through 6 presented quantitative analysis of published state appellate court decisions on free exercise disputes to answer the two major research questions at the heart of the dissertation: what factors predict claimant success in these decisions, and what factors predict the type of legal tests judges use in making their decisions? One of the most striking of these findings was the dramatically low frequency with which the courts sided with free exercise claimants. Only 12 percent of all cases at this level were decided in favor of the claimant. This low success rate differs from those reported in prior social scientific analyses (e.g., Brent 1999; Sisk et al. 2004; Claborn 2011), which were uniformly higher, though it is in line with the analysis of legal scholars (Ryan 1992; Lupu 1998). This may be due in part to the broader nature of what these researchers considered a free exercise case.
The key distinction between the success rate reported in this dissertation and in previous research, I believe, is that I restricted cases involving disputes over unique benefits that religious groups enjoy (e.g., tax-exempt status) from my analysis; when including such cases in my analysis, I observed a 33 percent claimant success rate, more consistent with previous social scientific analyses. An additional consideration is that my focus on state-level cases may also account for the low success rate, and that success rates in the federal courts may be systematically higher. Yet there is little reason to suspect this is true, at least during the cases settled between 1997 and 2006. In its 2006 decision in *Gonzales v. O Centro Espirita Beneficente Uniao do Vegetal* (546 U.S. 418), the Supreme Court recognized the applicability of the Religious Freedom Restoration Act (RFRA), passed by Congress in 1993 to require courts to use a particular form of legal test (the compelling interest test) which recognizes the validity of citizens’ religious as well as their legal obligations, in the federal court system. The Court had previously held in *City of Boerne v. Flores* that RFRA could not apply to the state courts, thus requiring the use of the more restrictive rational-basis test in the states for federal constitutional free exercise claims, but was silent on the law’s federal applicability. Claimants thus had little indication that they might have greater success in the federal courts, and as a result, I would not expect the success rates between the two court systems to be dramatically different as a result of claimants taking their cases to federal court and expecting a better result. Indeed, analysis by Brent (2003:564) of federal Courts of Appeals in the immediate years following the *Flores* decision suggests that claimants at this level had barely a 20 percent success rate, only somewhat greater than the success rate I observe at the state courts.

Another key descriptive finding of this study, consistent with earlier analysis (Wybraniec and Finke 2001), is that members of religious minority groups were disproportionately likely to
allege violations of their free exercise rights from 1997 to 2011. Members of new religious movements and Native American churches accounted for 10 percent of free exercise claims during this period despite accounting for about one percent of the U.S. population. Likewise, members of the most conservative Protestant sects accounted for nearly double the percentage of claims as their representation in the population would suggest if claims were made proportionately. This is entirely consistent with Donald Black’s (1976) contention that societal outsiders are more likely to find themselves out of favor with the law because their practices are less likely to have been legitimated through the lawmaking process. Practices that the majority approves of are more likely to be legalized or tolerated in society. In theory, marginal groups have a lesser degree of access to legal tools for legitimating their conduct. Thus, we see members of religious minority groups running afoul of the law in greater numbers than we would expect if the number of free exercise claims were proportional to the population.

Finally, I observed that more than half of all decisions in coded free exercise cases were made on the basis of this type of test, rather than the rational-basis one would expect to see more often at the state level given the Supreme Court’s ruling in City of Boerne v. Flores. As described in Chapter 2, the three factors leading to the use of a balancing test in these cases are:

(1) The passage of state-level versions of the federal Religious Freedom Restoration Act (RFRA) in about a third of all states, requiring that judges who rule on free exercise claims under state constitutional protections use a balancing test to do so, usually the compelling interest test articulated in the original RFRA;
(2) Congress’ enactment and the Supreme Court’s recognition of the Religious Land Use and Institutionalized Persons Act (RLUIPA) in 2000, which called for the use of a
balancing test in state cases involving the rights of the institutionalized and religious land use disputes; and

(3) Precedent in the state courts to interpret state constitutional free exercise protections using a balancing test.

The widespread use of a form of balancing test in state free exercise cases provides nationally representative evidence to support Durham and Smith’s (2010:94) claim that “a significant additional measure of free exercise protection has been re-established” in the wake of the Court’s Smith and Flores decisions. Though this is an imperfect measure of the diffusion of balancing tests nationally, relying as it does on an unequal number of cases from states with dramatically different populations, this finding demonstrates that within the distribution of free exercise cases heard at the state appellate level, use of the balancing test is widespread and very common.

Multivariate analysis in Chapters 5 and 6 allowed me to test sets of hypotheses based on legal positivist and legal realist logics about case outcomes and balancing test use. The dissertation makes two major findings regarding predictors of favorable decisions for claimants in these cases. First, I find that the most important predictor of favorable decisions is the form of legal test judges use: claimants were significantly more likely to receive favorable decisions when judges used balancing tests to evaluate their claims than when rational-basis tests were used. This relationship is highly significant, and when comparing the total effects of all predictors in Chapter 6, I found that it is by far the most powerful predictor of successes for claimants.

The primary policy implication of this finding is that advocates of the compelling interest test and other balancing test appear to be correct in arguing that the use of balancing tests expands the scope and depth of citizens’ free exercise rights. The analyzed data do not allow us
to draw conclusions about the efficacy of particular legislative efforts at the federal or state level in expanding these rights; the number of decisions in which judges cite a state-level Religious Freedom Restoration Act (a “mini-RFRA”) or the federal RLUIPA when their judgments is so small at the present time that even bivariate analysis of the effects of these forms of legislation would be futile. RLUIPA only applies to a subset of all free exercise claims, while not enough cases have reached the appellate courts after their state legislatures enacted mini-RFRAs (in states that have done so) to permit statistical analysis of any acceptable statistical power. Despite Claborn’s (2011:631) categorical claim that “state attempts to increase religious freedom do not work,” I do not believe that enough case law exists to allow us to draw generalizable, empirical conclusions on the effectiveness of either method at the present date. This analysis makes clear, however, that the mechanism by which legislative attempts to broaden the scope of free exercise rights – a form of balancing test, usually the compelling interest test – is effective at increasing the rates of claimant success.

The dissertation’s other major takeaway about case outcomes is a negative one: virtually no other hypothesized predictors were associated with differences in outcomes. This is particularly noteworthy because unlike previous studies (Way & Burt 1983; Brent 1999; Wybraniec & Finke 2001; Sisk et al. 2004), I did not find an association between claimants’ religious minority status and their likelihood of receiving favorable decisions. I confirmed this finding using three separate types of measures: a continuous measure of the state population proportion accounted for by the claimant’s religious classification, a series of dichotomous religious affiliation indicators and a single dichotomous measure of Christian/non-Christian status. In no instance did any of these measures predict case outcomes. As discussed in Chapter 5, the divergence in the results of this dissertation compared with prior studies on this measure
may be due to the exclusion of disputes over religious tax exemptions and other benefits given exclusively to religious groups. Claimants’ religious group size in the population is significantly and positively associated with favorable case outcomes in a model that includes all 453 cases originally coded, prior to the exclusion of this particular class of cases. Cases involving disputes over religious tax exemptions, which have higher rates of claimant success than the types of free exercise cases I included in my principal analysis, are more likely to involve larger, more “mainstream” religious groups; thus, the negative association between minority status and case outcomes observed in other studies may be due in part to the types of claims prevalent among these larger religions.

Yet while religious groups appear not to be treated differently by the state courts, the results of this study are not entirely positive for religious minorities; after all, the overall success rate in the free exercise cases studied in this analysis is still very low. Recall that minorities raise a disproportionate number of free exercise claims relative to their representation in the U.S. population. The fact that all religious groups fare poorly when pursuing free exercise claims in the courts means that claims by religious minorities are consistently ruled against. This may have the effect expected by legal scholars and religious economies theorists alike that these groups are at a greater risk of being severely restricted in their ability to practice their religions than are more mainstream religions, potentially to the point of being unable to engage in religious practice altogether. The results of this study highlight the tenuousness of minority religious groups’ ability to operate when they run afoul of generally applicable, religiously neutral laws, and this is true regardless of the form of legal test judges employ in adjudicating their claims of free exercise violations.
In addition to these findings about case outcomes, I can also draw conclusions based on my analysis about the factors that influence whether judges will apply a balancing test or a rational-basis test to the facts of each case. To my knowledge, this is the first study of free exercise cases that models the predictors of legal test use. Moreover, I observed that a number of hypotheses based on the premise that judges consider more than just precedent in choosing legal tests were supported. The models in Chapter 6 found modest support for hypothesized relationships based on judges’ characteristics and the structural constraints in which they operate – hypotheses grounded in the realist perspective – regarding the choice of the balancing test as the standard for evaluating claims. For instance, Hypothesis Five-A, regarding the negative association between the Republican composition of the bench and the likelihood that a balancing test will be used, was supported in Chapter 6. In addition, I found evidence that differences in the ways judges retain their seats are associated with the legal tests they select to evaluate cases. Judges who require public approval to retain their seats at the end of their term appear more likely to employ balancing tests in evaluating free exercise claims net of all other factors. However, this conclusion should be tempered somewhat, since this effect is only observed when controlling for an additional variable – the importance of religion in the lives of people in each state – which is substantially correlated with one of the judicial selection system indicators.

The results of models testing these legal realist hypotheses about an area where judges have some degree of agency – the standard they use in guiding how they will evaluate claims – offer a good deal of support for what judicial behavioralists and other legal realists have argued: when studying judges’ actions in an empirical manner, it is critical that judges’ characteristics and the characteristics of the environment in which they operate be taken into account. This is particularly true in light of the fact that in all models testing these hypotheses, the role of
precedent – a quintessential positivist explanation for explaining judicial action – was controlled for.

**Limitations of Study and Directions for Further Research**

Though the findings reported in this dissertation have much to offer to those involved in policy debates over religious free exercise, as well as scholars in a number of sociological sub-disciplines, the study has a number of limitations that should be recognized. First, as described in Chapter 3, these results cannot and should not be generalized to all state-level free exercise cases. There are two reasons for this. First, the analysis was only conducted on state appellate court opinions that have been released for publication, and necessarily excludes opinions that have not been published. While the LoislawConnect database used to locate opinions does contain unpublished opinions for a small number of states, the lack of access to unpublished opinions from the majority of states led me to exclude unpublished opinions from the analysis. Scholars disagree on whether to include unpublished opinions in their analysis (for instance, James Brent (1999, 2003) elected to do so in his analysis of free exercise decisions), but consistent with the logic of Sisk et al. (2004), I excluded the unpublished decisions on the grounds that including the limited number of available decisions of this type could bias the results of my analysis.

Second, the cases analyzed here were drawn from the appellate levels of state court; trial court opinions were not considered here, since they were not available in the vast majority of states through primary source database services like LoislawConnect. This limits the pool of cases to only those that were appealed by one party or another; cases that were heard at the initial trial level but not at the appellate level are not included in this analysis. The dual limitations of
the pool of cases to (a) only those that were published and (b) only those that had been appealed at least once means that it would be an error to generalize these findings to all state free exercise claims. Research incorporating trial court opinions into the study’s analytic framework would greatly enhance the generalizability of its findings.

The study is also limited in the number of cases available for analysis. Only 225 state appellate free exercise cases met the criteria for inclusion in the analysis; while conclusions may be drawn from regression analyses using this number of observations, the low level of variation on the primary outcome variable (cases outcomes) and the number of parameters that had to be estimated increases the possibility of Type II error. This is evident in a number of tables, where large effect sizes were not noted as describing significant associations due to even larger standard errors. The relatively small number of observations in the analysis, while limiting the power of the analysis to identify effects as statistically significant, does highlight the size and power of the effects that are noted as significant (particularly with greater than 95 percent confidence). Nevertheless, the addition of a greater number of cases would add enormously to the ability of the models estimated in these analyses to test hypotheses accurately.

One aspect of the study that is both a limitation of the present study and a clarion call for future research is that federal appellate cases during the same period of time were not coded for this analysis. While cases in the federal court system were subject to different sets of constraints (e.g., the Court’s Flores decision did not apply explicitly to cases in federal courts, and the effects of judicial retention systems cannot be tested in the same way), it would be highly illustrative to compare cases of this type to state-level cases on a variety of dimensions. Differences between the two court systems regarding the variables that are significantly associated with favorable outcomes for claimants could be assessed and compared. In addition,
uncertainty in the federal courts over whether the compelling interest test mandated by the Religious Freedom Restoration Act should be applied in the wake of *Flores* created a situation of tremendous opportunity for researchers interested in judicial behavior in absence of clear precedents. Content analysis of federal appellate courts akin to the effort conducted for this dissertation and analysis of the resultant data would lead to a broader opportunity to test the validity of this dissertation’s results in a wider setting.

Another avenue for further inquiry may be to explore and explain arguably the most unusual finding of this dissertation: the very strong and negative association between the importance of religion for state populations and the use of balancing tests in those states for evaluating free exercise claims reported in Chapter 6. Why might the rational-basis test, the use of which erects substantially higher barriers to religious free exercise in these situations, be so much more common in states where religion appears to be more important at the population level? Perhaps there is an association between religious salience and political conservatism which explains this finding; after all, the likelihood that a balancing test would be used decreased as the proportion of Republican judges who heard a given case increased. Controlling for the latter variable would not account for states’ political leanings, as large proportions of Republican judges were to be found in heavily Democratic states, and vice versa. Future research might explore this curious connection and determine whether broader political factors account for this linkage.

This dissertation explored the state of religious free exercise in America through an examination of its most uncertain terrain: how it is defined when it comes into conflict with the institutions of the law. The results of the study indicate that despite attempts to broaden its meaning in these circumstances, judges have a far greater inclination to hold that citizens’
religious obligations are secondary to their obligations under the law of the land. Nevertheless, as the period from the Supreme Court’s 1990 decision in Employment Division v. Smith to the present day demonstrates, the boundaries of the right to the free exercise of religion remain in flux, and further judicial and legal action may shift those boundaries in unexpected ways. Legal scholars and social scientists alike would be well-served by keeping a watchful eye for shifts in these boundaries, for as the American population grows more religiously diverse, the status quo is never likely to remain the status quo for long.
Appendix A: Coding Instrument for Content Analysis

CODEBOOK

1) STATEID
   This is the case ID. For each case you code, the “StateID” is the postal abbreviation and the order in which it appears when searching. For example, in Alaska, “Sands v. Living Word Fellowship” should be the #6 case you see. If you were to code this case, its “StateID” value would be AK6.

2) NAME
   Case name. If you coded “Sands v. Living Word Fellowship,” you would enter that into this field (“Sands v. Living Word Fellowship”, with no quotes).

3) CITE
   Case citation. This is usually right next to the case name; for example, Sands v. Living Word Fellowship’s case citation is “34 P.3d 955 (Alaska 2001)”. If it is not given, use the docket number (see below)

4) DOCKET
   Case docket number. This is usually below the names of all the parties to the case. For Sands, the docket number is “S-9031”.

5) DATE
   Date that the opinion was issued. Format is like this: YYYYMMDD. For example, September 30, 2012 would be 20120930.

6) NOCODE
   Reason why the free exercise case wasn’t coded. (You can just skip over most cases, but there are some cases where someone is making a free exercise claim that you won’t code. If you code anything other than a 0, stop here. Some reasons for not coding cases:
   (0) You DID code this case.
   (1) Intrareligious dispute. (Opposite sides are members of the same religious group, each making a free exercise claim. You rarely code cases like this, unless one side of the group is NOT making a free exercise claim (for example, someone sues for workplace discrimination, and the other side makes a free exercise claim for why they should be allowed to discriminate).
   (2) Outside period of analysis. (Case was heard before June 25, 1997.)
   (3) Free exercise claim was made, but not ruled upon. (Sometimes judges will throw out a case without even ruling on the free exercise claim; this will happen when the case is about more than just religion.)
   (4) Pure Establishment Clause case. (The case is decided based on the constitution’s Establishment Clause; these are “separation of church and state” cases.)
   (5) Unusual case. (The case is about something where no sane judge would ever consider the free exercise case. For example, someone saying that their free exercise rights let them kill or rape someone.)
   (6) Hypothetical claim. (The case is about a free exercise violation that might happen, but hasn’t happened yet. You WOULD code hypothetical claims in child custody cases, but that is the only exception.
   (7) Case remanded to lower court. (Often, judges will decide to remand a case, or send it back down to a lower court to be heard. This often happens when judges decide that the lower court didn’t decide it using the right legal test. For example, a higher court may decide that the lower court should’ve used a balancing test. We don’t code these cases because there’s no decision yet.)
   (8) Procedural. (Judge makes a ruling based on some procedural issue rather than considering the arguments. For example, they throw out the case because the claimant didn’t make their free exercise claim in the lower court or trial court.)
   (9) Private matter. (This is rare, but there may be some employment discrimination claims based on religion. We don’t code these unless the employer is a government body. You would code unemployment compensation claims though.)
   (OTHER) Some other reason. Write the other reason in this field.
7) STATE
Two-letter state postal code.

8) LEVEL
Level of court. The name of the court is given near the top of the opinion (e.g., “Supreme Court of Alaska”). Use the charts at http://www.courtstatistics.org/Other-Pages/State_Court_Structure_Charts.aspx as a guideline.
   1 = Intermediate Appellate Court
   2 = Court of Last Resort

9) WIN
Did the free exercise claimant get a favorable decision? (Code “Yes” if they got anything they wanted, not including having their case heard again somewhere.)
   0 = No
   1 = Yes

10) APPELLN
Is the free exercise claimant appealing the decision of the lower court?
   0 = No
   1 = Yes

11) RELIG
Religious affiliation of the free exercise claimant. Be as specific as you can based on what’s actually given in the text of the opinion.

12) SHERBERT
Is Sherbert v. Verner (or just Sherbert) cited at all in the opinion? [Note that for these citation questions, opinions may use longer or shorter versions of the names of the cases. Code “Yes” if any version of these case names are cited.]
   0 = No
   1 = Yes

13) YODER
Is Wisconsin v. Yoder cited at all in the opinion?
   0 = No
   1 = Yes

14) SMITH
Is Employment Division, Department of Human Resources v. Smith (or just Smith) cited at all in the opinion?
   0 = No
   1 = Yes

15) OCENTRO
Is Gonzales v. O Centro Espirita Beneficente Uniao do Vegetal cited at all in the opinion?
   0 = No
   1 = Yes

16) STATUTE1
Does the free exercise claimant make their claim based on some legal statute? (For example, on RFRA, RLUIPA, or a state “mini-RFRA”?)
   0 = No
   1 = Yes

17) STATUTE2
If STATUTE=1, what is the name of the statute? (Write it out.)

18) MINIRFRA
Did the claimant make their claim based on a state-level RFRA?

0 = No
1 = Yes

19) TESTNAME
What is the name of the court opinion or legal statute on which the judges made their ruling about the free exercise claim? (Write it out.)

20) TEST1
What type of legal test did the judges use to make their ruling on the free exercise case?

1 = Employment Division v. Smith, or some other “rational-basis” test that says that as long as a law is rationally related to a government interest and/or it doesn’t discriminate on religion, there are no exceptions to the law. (Examples include Turner v. Safley, O’Lone v. Estate of Shabazz, Church of the Lukumi Babalu Aye v. City of Hialeah, and others.)

2 = A “balancing test.” These are tests that take burdens on religious rights into consideration. (Major examples are Sherbert v. Verner, Wisconsin v. Yoder, RFRA, RLUIPA, and state-level RFRAs. There are also other state cases. Tests like this may be called “compelling interest tests” or “strict scrutiny.”)

3 = Something other than a rational-basis test or a balancing test. Might be based purely on facts, might be based on tax law, etc.

21) RFRA
Is the federal Religious Freedom Restoration Act or RFRA cited at all in the case?

0 = No
1 = Yes

22) RFRASTAT
If RFRA = Yes, how is the status of the federal RFRA interpreted?

1 = RFRA was overturned (no more specifics than that)
2 = RFRA is not applicable to the states
3 = RFRA was cited as if it still applies
4 = RFRA is cited, but no specifics about its status are given
99 = RFRA is not cited at all

23) BOERNE
Is City of Boerne v. Flores cited at all in the opinion?

24) COMPEL
Does the court cite a possible compelling state interest in making its ruling, even if it doesn’t agree that the interest in compelling?

0 = No
1 = Yes, and the court agrees that the interest is compelling
2 = Yes, but the court doesn’t agree that the interest is compelling

25) INTEREST
If COMPEL = 1 or 2, what government/state interest or interests are cited? (Code this as “99” if COMPEL = 0)

26) HYBRID
Was it mentioned that the free exercise claimant or court said this was a “hybrid” case? (“Hybrid” free exercise claims are a certain type of claims described in the Smith opinion that are combinations of free exercise claims and some other type of rights claim, like freedom of speech.)

0 = No
1 = Yes

27) HYBRID2
If HYBRID=YES, did the court agree that this was a “hybrid” case?

0 = No
Note on the TYPE_____ Variables: Only code one of these as a YES.

28) TYPEINST
Type of case: Was this a case involving the rights of an institutionalized person (e.g., a prisoner)?
   0 = No
   1 = Yes

29) TYPELAND
Type of case: Was this a land use case?
   0 = No
   1 = Yes

30) TYPEUNEM
Type of case: Was this an unemployment compensation case?
   0 = No
   1 = Yes

31) TYPEJOB
Type of case: Was this some other type of employment-related case? (Note: don’t code private employment discrimination cases at all, unless there’s some government involvement like there is in an unemployment comp case.)
   0 = No
   1 = Yes

32) TYPETAX
Type of case: Was this a tax-related case? (For example, does the claimant want a religious tax exemption?)
   0 = No
   1 = Yes

33) TYPEIMM
Type of case: Was this a case regarding immunizations or vaccination?
   0 = No
   1 = Yes

34) TYPEID
Type of case: Was this a case involving a form of government-issued ID? (for example, something regarding Social Security, a driver’s license, or a passport)
   0 = No
   1 = Yes

35) TYPEEDUC
Type of case: Was this case related to education?
   0 = No
   1 = Yes

36) TYPEGOVT
Type of case: Was this case related to government operations in ways not covered in any other case type above?
   0 = No
   1 = Yes

37) TYPEDRUG
Type of case: Was this a drug-related case, in a way not covered by any other case type above?
   0 = No
   1 = Yes
38) TYPEKID
Type of case: Was this a child custody or visitation case?
0 = No
1 = Yes

39) TYPECRIM
Type of case: Was this some other type of criminal case not covered by the categories above?
0 = No
1 = Yes

40) TYPEOTHR
Type of case: Did this case involve some other issue other the types given above?
0 = No
1 = Yes

41) OTHER
If TYPEOTHR = YES, describe the substantive issue of the case.

42) NUMBER
Size of the free exercise claimant’s group.
1 = Individual claimant
2 = Number of individual claims not unified as a formal organization
3 = Autonomous organization

43) LITERAL1
Are the words “Free Exercise Clause” or “free exercise of religion” (or a close variant of “free exercise of religion”) contained in the opinion?
0 = No
1 = Yes

44) LITERAL2
If LITERAL1 = 0, Does the claimant make their case based on “freedom of religion,” “religious freedom,” or something vague like those terms?
0 = No
1 = Yes
99 = LITERAL1 was coded as YES.

45) STCONST
Does the free exercise claimant make their claim based on the state constitution or a state RFRA?
0 = No
1 = State constitution (that is, the state constitution’s free exercise clause)
2 = State RFRA
3 = Both the state constitution and a state RFRA

46) FCONST
Does the free exercise claimant make their claim based on the federal constitution?
0 = No
1 = Yes, on “constitutional rights” or something vague without specifically indicating the federal constitution
2 = First Amendment, Free Exercise Clause, or federal constitution ARE referenced directly

47) DISSENT
Does at least one of the judges dissent from the majority opinion on the free exercise issue?
0 = No
1 = Yes
NOTE ON JUDGE VARIABLES: If the opinion indicates that the judge is the “presiding” judge or the Chief Justice, write “(P)” before their name. If the judge dissented from the majority opinion on the free exercise issue, write “(D)” before their name.

48) JUDGEW
Last name of judge who wrote the majority opinion. If “per curiam” (which means all the judges signed off on the opinion), write “per curiam.”

49) JUDGE1, JUDGE2, JUDGE3, etc.
Write the last name of the other judges given in this opinion. Write the names of the judges who dissented on the free exercise case last.
### Appendix B: Distributions of Cases by Values of Dependent Variables

<table>
<thead>
<tr>
<th>Nominal/Dichotomous Measures (Expressed as Number of Decisions):</th>
<th>Outcome of Free Exercise Claim:</th>
<th>Form of Legal Test Used:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal Test Used to Evaluate Claim</td>
<td>Upheld</td>
<td>Rejected</td>
</tr>
<tr>
<td>Balancing Test</td>
<td>21</td>
<td>105</td>
</tr>
<tr>
<td>Rational-Basis Test</td>
<td>5</td>
<td>94</td>
</tr>
</tbody>
</table>

#### Case Type

| | Prisoners’ Rights/Land Use Cases | Education/Child Welfare Disputes | Disputes over Positive Gov’t. Policies | Other Disputes |
| | 14 | 67 | 45 | 36 |
| | 4 | 40 | 21 | 12 |
| | 3 | 30 | 26 | 18 |
| | 5 | 62 | 34 | 33 |

#### Judicial Retention System

| | Retention Election | Competitive Election | No Direct Election |
| | 7 | 79 | 48 | 38 |
| | 12 | 70 | 49 | 33 |
| | 7 | 50 | 29 | 28 |

#### Precedent for Balancing Test Use

| Precedent | 21 | 114 | 84 | 51 |
| No Precedent | 5 | 85 | 42 | 48 |

#### Continuous Measures (Expressed as Means):

| Proportion of GOP Judges | .530 | .443 | .425 | .491 |
| Statewide Importance of Religion | .520 | .540 | .523 | .555 |

#### Interaction: Religious Group Size and Judicial Retention System

| | Religious Group Size x Retention Election | Religious Group Size x Competitive Election | Religious Group Size x No Election* |
| | -.691 | -1.172 | -1.133 | -1.254 |
| | -1.299 | -1.171 | -1.043 | -1.107 |
| | -1.321 | -.632 | -.610 | -.913 |

\(N = 225.\) Means and frequencies are derived from imputation dataset \(m = 1\), with the exception of the starred interaction term “Religious Group Size x No Election.” Values of this interaction term were not imputed, and means are derived from listwise deletion.
Appendix C: Path Diagram of Structural Equation Model Predicting Balancing Test Use and Favorable Outcomes for Claimants, 1997-2011

Note: Unstandardized probit regression coefficients are reported. *** p < .001; ** p < .01; * p < .05; † p < .10. N = 225 using 20 imputations. Correlations between exogenous variables are accounted for but not shown.

Model fit: $\chi^2 = 10.476$ (1 df); CFI = 1.000; RMSEA = 0.000
Appendix D: Path Diagram of Trimmed Structural Equation Model Predicting Balancing Test Use and Favorable Outcomes for Claimants, 1997-2011

Note: Unstandardized probit regression coefficients are reported.*** p<.001; ** p<.01; * p<.05; † p<.10. N = 225 using 20 imputations. Correlations between exogenous variables are accounted for but not shown.

Model fit: $\chi^2 = 10.445$ (13 df); CFI = 1.000; RMSEA = 0.000
Appendix E: Estimates from Structural Equation Model Predicting Balancing Test Use and Favorable Outcomes for Claimants, 2000-2011

<table>
<thead>
<tr>
<th>Variable</th>
<th>Balancing Test</th>
<th>Favorable Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precedent for Balancing Test Use</td>
<td>-.731</td>
<td>----</td>
</tr>
<tr>
<td>Religious Group Size (Proportion of State Population, Logit)</td>
<td>-.002</td>
<td>-.039</td>
</tr>
<tr>
<td><strong>Case Type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/Child Welfare Disputes</td>
<td>-.121</td>
<td>-.425</td>
</tr>
<tr>
<td>Disputes over Positive Gov’t. Policies</td>
<td>.350</td>
<td>-.697</td>
</tr>
<tr>
<td>Other Disputes</td>
<td>-.280</td>
<td>-.316</td>
</tr>
<tr>
<td><strong>Metro Area Minority Pop Size (Log)</strong></td>
<td>.112†</td>
<td>.009</td>
</tr>
<tr>
<td><strong>Proportion of GOP Judges</strong></td>
<td>-.731*</td>
<td>.651</td>
</tr>
<tr>
<td><strong>Judicial Retention System</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>.651*</td>
<td>-.331</td>
</tr>
<tr>
<td>Competitive Election</td>
<td>.994**</td>
<td>-.191</td>
</tr>
<tr>
<td><strong>Statewide Importance of Religion</strong></td>
<td>-4.654***</td>
<td>.967</td>
</tr>
<tr>
<td><strong>Use of Balancing Test</strong></td>
<td>----</td>
<td>.442**</td>
</tr>
<tr>
<td>McKelvey and Zavoina Pseudo-$R^2$</td>
<td>.200</td>
<td>.281</td>
</tr>
</tbody>
</table>

*** $p<.001$; ** $p<.01$; * $p<.05$; † $p<.10$. N = 178 using 20 imputations. Unstandardized probit regression coefficients are reported.

Model Fit: $\chi^2 = .077$ (df = 1); RMSEA = .000; CFI = 1.000.

Reference Categories: **Case Type:** Prisoners’ Rights/Land Use Disputes. **Judicial Retention System:** No Direct Election.
## Appendix F: Polychoric Correlations Between Analyzed Variables

<table>
<thead>
<tr>
<th></th>
<th>Outcome of Case</th>
<th>Use of Balancing Test</th>
<th>Religious Grp. Size</th>
<th>Bal. Test Precedent</th>
<th>Prop. GOP Judges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of Balancing Test</td>
<td>.393</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious Group Size (Logit)</td>
<td>-.125</td>
<td>.100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balancing Test Precedent</td>
<td>.252</td>
<td>.194</td>
<td>-.112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion GOP Judges</td>
<td>.131</td>
<td>-.107</td>
<td>-.039</td>
<td>.095</td>
<td></td>
</tr>
<tr>
<td>Statewide Imp. of Religion</td>
<td>-.106</td>
<td>-.191</td>
<td>-.100</td>
<td>-.099</td>
<td>-.105</td>
</tr>
<tr>
<td>Metro Area Minority</td>
<td>.000</td>
<td>.107</td>
<td>.121</td>
<td>.032</td>
<td>.226</td>
</tr>
<tr>
<td>Population Size (Log)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educ./Child Welfare</td>
<td>-.066</td>
<td>.039</td>
<td>.144</td>
<td>-.215</td>
<td>-.133</td>
</tr>
<tr>
<td>Pos. Gov’t. Policies</td>
<td>-.056</td>
<td>.081</td>
<td>.004</td>
<td>-.072</td>
<td>.094</td>
</tr>
<tr>
<td>Other Disputes</td>
<td>-.149</td>
<td>-.086</td>
<td>-.006</td>
<td>-.103</td>
<td>.014</td>
</tr>
<tr>
<td>Judicial Retention System</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>-.144</td>
<td>-.004</td>
<td>.000</td>
<td>-.049</td>
<td>.077</td>
</tr>
<tr>
<td>Competitive Election</td>
<td>.115</td>
<td>.072</td>
<td>-.022</td>
<td>.109</td>
<td>-.112</td>
</tr>
</tbody>
</table>

(continued on next page)
### Appendix F: Polychoric Correlations Between Analyzed Variables (continued)

<table>
<thead>
<tr>
<th></th>
<th>Statewide Imp. of Religion</th>
<th>Metro Minority</th>
<th>Case Type: Educ./Child Welfare</th>
<th>Case Type: Pos. Gov’t Policies</th>
<th>Case Type: Other Disputes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statewide Imp. of Religion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metro Area Minority</td>
<td>-.345</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population Size (Log)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educ./Child Welfare</td>
<td>.118</td>
<td>-.101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pos. Gov’t. Policies</td>
<td>-.028</td>
<td>.063</td>
<td>-.204</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Disputes</td>
<td>-.061</td>
<td>-.031</td>
<td>-.321</td>
<td>-.270</td>
<td></td>
</tr>
<tr>
<td>Judicial Retention System</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention Election</td>
<td>.008</td>
<td>-.065</td>
<td>-.111</td>
<td>.036</td>
<td>.108</td>
</tr>
<tr>
<td>Competitive Election</td>
<td>.470</td>
<td>-.241</td>
<td>.139</td>
<td>-.079</td>
<td>-.089</td>
</tr>
</tbody>
</table>

### Judicial Retention System:
- Retention Election
  - .596
- Competitive Election
  - .596
REFERENCES:

Works Cited:


Cases Cited:

*Alexander v. State*, 282 S.W.3d 143 (Texas 2009)


*Austin v. Crosby*, 866 So.2d 204 (Florida.App. 5 Dist. 2004)

*Barr v. City of Sinton*, 295 S.W.3d 334 (Texas Court of Appeals 2005)

*Barr v. City of Sinton*, 295 S.W.3d 287 (Texas 2009)

*Cannon v. Mote*, 824 N.E.2d 1227 (Illinois 2005)

*Cantwell v. Connecticut*, 310 U.S. 296 (1940)


*City of Boerne v. Flores*, 521 U.S. 507 (1997)

*Conley v. Roman Catholic Archbishop*, 102 Cal.Rptr.2d 679 (California 2000)


*Humphrey v. Lane*, 728 N.E.2d 1039 (Ohio 2000)


*Malicki v. Doe*, 814 So.2d 347 (Florida 2002)

Meggett v. PA Dept. of Corrections, 892 A.2d 872 (Pennsylvania 2006)


Reid v. Arkansas, 2011 Ark. 184 (Arkansas 2011)

Reynolds v. United States, 98 U.S. 145 (1878)

Shepherd Montessori Center v. Ann Arbor Township, 783 N.W.2d 695 (Michigan 2010)

Sherbert v. Verner, 374 U.S. 398 (1963)


Spell v. Muhammad, 756 So.2d 748 (Mississippi 2000)

State v. Miller, 549 N.W.2d 235 (Wisconsin 1996)


Thompson v. Department of Corrections, 18 P.3d 1198 (California 2001)


Wisconsin v. Yoder, 406 U.S. 205 (1972)
Robert R. Martin

Department of Sociology & Crime, Law and Justice
The Pennsylvania State University
211 Oswald Tower
University Park, PA 16802

Office Phone: (814) 865-6258
Fax: (814) 863-7216
Cell Phone: (814) 954-0180
E-mail: robert.martin@psu.edu

Education
Ph.D., Sociology, The Pennsylvania State University, August 2013
  Exercise Claims in the State Appellate Courts, 1997-2011.”
  Committee: Roger Finke (chair), Duane Alwin, John McCarthy, Christopher Zorn
  Major Area of Specialization: Sociology of Religion
  Minor Areas of Specialization: Quantitative Methods, Sociology of Organizations

M.A., Sociology, Fordham University, May 2009
  Master’s Thesis: “The Effect of Believing in an Involved God on Volunteering Likelihood.”
  Committee: Evelyn Bush (chair), Matthew Weinshenker

B.F.A., Film and Television Production, Minor in Sociology, New York University, January 2007

Publications
  Quantitative Assessment of Free Exercise Cases in the U.S. State Courts, 1981-2011.” In Religious
  Freedom in America: Constitutional Traditions and New Horizons, ed. Allen D. Hertzke. Norman, OK:
  University of Oklahoma Press.

Reports

Professional Positions
2011-2012 Instructor, Department of Sociology & Crime, Law and Justice, The
  Pennsylvania State University

2009-2013 Research Associate, The Association of Religion Data Archives
  (www.theARDA.com), Department of Sociology & Crime, Law and Justice,
  The Pennsylvania State University

Teaching Experience
2012 Instructor, Introductory Social Psychology (The Pennsylvania State University)

2011 Instructor, Social Problems (The Pennsylvania State University)