AUTOMATED CONTENT ANALYSIS AND THE DEVELOPMENT AND UTILIZATION OF LEGAL DOCTRINE IN THE FEDERAL COURTS

by

CHASE MARTIN PORTER

JOSEPH L. SMITH, COMMITTEE CHAIR
STEPHEN BORRELLI
THOMAS G. HANSFORD
RICHARD C. FORDING
DANA PATTON

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Political Science
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2018

ABSTRACT

The citation and interpretation of precedent and the development and utilization of legal doctrine are distinct concepts. Whereas previous literature has focused on the use of precedent, this study makes a theoretical argument for the importance of distinguishing between precedent and doctrine and applies automated content analysis tools to the measurement of legal doctrine in court opinions. These tools are used to study doctrinal utilization by the Supreme Court and the circuit courts in the United States judicial system.

From a theoretical perspective, this study leverages a qualitative case study of the development and application of the *Lemon* test in Establishment Clause jurisprudence to illustrate the importance of carefully distinguishing between the concepts of precedent and doctrine. The case study exposes potential weaknesses in dependence upon *Shepard's Citations* as a tool for understanding the development of legal doctrine. The concept of doctrinal vitality is proposed as a way to measure the impact of legal doctrine across time.

Given the difficulties that are inherent in measuring a qualitative concept (language) quantitatively, a careful examination of various automated text analysis methodologies was conducted. The programming language Python was used to analyze the doctrinal composition of court opinions through unsupervised topic modeling and supervised sentence counting. Utilization of doctrinal language was modeled as a function of variables that impact judicial behavior on the Supreme Court and circuit courts, including doctrine age, judicial ideology, doctrinal vitality, opinion characteristics, and hierarchical effects. While the substantive findings

are mixed, this dissertation considers important theoretical implications regarding the development and utilization of legal doctrine and explores the potential benefits and challenges related to the use of automated text analysis in the study of legal doctrine.

DEDICATION

To Natalie: All our wishes came true, didn't they? Thank you – for everything. To Chase Jr.: Remember who you are – you are my son. To Mary Beth: I can't wait to be part of your world. I love you all. We did it. Let's go to Disneyland.

LIST OF ABBREVIATIONS AND SYMBOLS

β	Regression coefficient
LR	Likelihood ratio
N	Sample size
p	Probability of obtaining a value as extreme as the sample if the null hypothesis is true
r	Pearson's correlation coefficient
SE	Standard error
t	Ratio of the difference between population means relative to the standard error
X^2	Measure of the goodness of fit between observed data and expected data
=	Equal to
<=	Less than or equal to

ACKNOWLEDGMENTS

Graduate students sometimes struggle with the question of how to find a faculty mentor. Looking back on my journey at Alabama, it was always obvious who my mentor was going to be, even before I began my time in the program. From the moment I visited the department as a prospective student, Jos Smith took a deep interest in my academic interests and in me as a person. Frankly, when I started my Ph.D., I had no idea what I was doing and no clear concept of what I wanted to study. All I knew was that I was really interested in law, politics, and religion. Over the past five years, Joe has nurtured those interests while offering countless hours of mentorship in every aspect of becoming a political scientist. I simply cannot offer enough gratitude for what he has done: chairing this dissertation, working on research together, training me in a programming language, preparing me for conferences and the job market, writing recommendations. Thank you, Joe. I anticipate many more years of friendship.

I am grateful for the other members of my dissertation committee. Tom Hansford co-authored the book that inspired this dissertation. After a chance meeting at the annual meeting of APSA, he graciously agreed to review what I was doing with his work and to serve as the external member of my committee. Admittedly, it is an intimidating proposition to have an accomplished scholar who inspired your work evaluate that very work, but he has proven to be a great encourager who challenged me to sharpen the work in several important ways (not the least of which was strengthening the entire theoretical foundation). In addition to lots of job market advice (and lots of recommendations), Rich Fording strengthened this dissertation with specific methodological advice that improved my models. Steve Borrelli encouraged me to do more

qualitative work to help bring out the nuance of the legal areas that I studied. Dana Patton graciously agreed to join the project at a later stage and pushed me to the finish line with the consistent reminder that "a good dissertation is a finished dissertation." This dissertation is finished; I can only hope that it is also good. I am also grateful for her advice and support on the job market, especially in preparation for the flyout interview that is the reason I have a job next year. Allen Linken provided extremely helpful feedback on the dissertation prospectus. Some of his ideas, including the importance of considering individual justice effects, made it into the final dissertation; others were sacrificed on the altar of a good dissertation being a finished dissertation, and are part of the future research agenda. Allen also provided a lot of recommendations and mentoring; it is a privilege to count him as a friend.

Many other scholars provided helpful feedback on various phases of this project. At the risk of omitting someone, several come to mind. Amanda Bryan's enthusiasm for the use of text analysis to study doctrinal development in the first iteration of this project encouraged me to continue with this research. H.W. Perry helped me see the need to delineate and develop the several different aspects of the project that were originally a bit muddled. Kelly Rader offered helpful challenges to the early focus on jurisprudential regimes and helped me reorient the project in more fruitful theoretical directions. Robert Hume helped me to clarify how I operationalized key measures such as doctrinal age. Rachel Hinkle offered important insights on how to conduct automated text analysis. Benjamin Kassow's and Maureen Stobb's enthusiasm for the work was a great encouragement, and their insights on clarifying how I framed, articulated, and emphasized the theoretical contributions of the project helped plant the seeds for a greatly improved theory chapter. Deborah Beim and Lauren Bell pushed me to answer the "why does this project matter" question much more clearly than I originally had answered it.

Susan Haire emphasized the need to demonstrate the face validity of my utilization measures and offered helpful suggestions for considering other areas of law besides the ones I originally studied. Matthew Hitt challenged me to clarify my critique of *Shepard's Citations* relative to its use in the political science literature and to think about the implications of the uniqueness and ambiguity of the *Lemon* test. All of these scholars offered many more incredibly useful insights that sharpened the project at various stages; the improvements are theirs, the continued room for improvement is mine. To anyone who I have failed to mention who offered feedback on parts of this project at the 2016, 2017, or 2018 meetings of MPSA, the 2016 UT-Austin Graduate Conference in Public Law, the 2017 or 2018 meetings of SPSA, or the 2017 meeting of APSA, I am grateful for the ways in which you challenged and shaped my thinking.

While I could not begin to list all of the ways in which my fellow graduate students at Alabama have impacted this work and my life, I am grateful for each one of you. You know who you are and what you have done. I have been in some form of school for approximately twenty-six of my thirty-one years on this earth; my parents have been a constant system of support for every one of them. I owe them a debt of gratitude I simply can never repay; moreover, I owe them my life. Mom and Dad, I'm finally done with school…I think. My entire time at Alabama has been a joint effort with my wife Natalie; this dissertation belongs to her as much as it does to me. To all of the faculty at Alabama who were not directly involved with this project but have taught me, mentored me, or encouraged me in some way: thank you. And to the faculty and administration of California Baptist University, who took a chance on an ABD and thus allowed me to achieve a dream I have had for many years: I am eternally grateful and eagerly await many fruitful years as your colleague and friend.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

In the 2008 case *United States* v. *Williams* (553 U.S. 285), a 7-2 Supreme Court reversed a ruling from the Eleventh Circuit that held that a federal statute prohibiting the possession and distribution of child pornography was overbroad and vague. Writing on behalf of the majority, Justice Scalia rejected the argument that the Court needed to subject a "content-based restriction" to "strict scrutiny," quoting from the Eleventh Circuit opinion in the process (*United States* v. *Williams* 2006, 1298). The Eleventh Circuit attributed this idea to the Supreme Court in *United States* v. *Playboy* (2000, 529 U.S. 803), which grounded the doctrine in *Sable Communications of California* v. *FCC* (1989, 429 U.S. 115). In *Sable*, the Court noted that "the Government may, however, regulate the content of constitutionally protected speech in order to promote a compelling interest if it chooses the least restrictive means to further the articulated interest" (*Sable Communications of California* v. *FCC* 1989, 126). Continuing the chain of influence, the *Sable* Court grounded this idea in three previous Court opinions, which each contained a curious feature: although they were decided after *Grayned* v. *City of Rockford* (1972, 408 U.S. 104), none of them linked this content-regulation doctrine to *Grayned*. Yet, the *Grayned* Court argued that "our cases make clear that, in assessing the reasonableness of a regulation, we must weigh heavily the fact that communication is involved; the regulation must be narrowly tailored to further the State's legitimate interest" (*Grayned* v. *City of Rockford* 1972, 116-17).

In *Playboy*, the Court noted that "if a statute regulates speech based on its content, it must be narrowly tailored to promote a compelling Government interest" (*United States* v.

Playboy 2000, 813). Compare the quote to *Grayned*: "further" becomes "promote," "State" becomes "Government," and "legitimate" becomes "compelling." And yet, the *Playboy* Court never cited *Grayned*. Nor did the *Sable* Court. The Eleventh Circuit cited *Grayned* in *Williams*, but not for the content regulation doctrine. Somehow, the *Playboy* Court essentially directly lifted language from a case decided twenty-nine years prior, built a river of precedent to justify the use of that doctrinal language, and then never acknowledged the original headwaters. Perhaps doctrinal development on the Court resembles a game of telephone where the message undergoes minor modifications and the original speaker fades into the background.

A better way of understanding the Court's game of telephone may be to see the game as a conference call, where there are so many original speakers that citing the "correct" one becomes impossible. Although the *Playboy* Court did not cite *Grayned*, the Court did discuss the doctrine of content neutrality. It grounded the doctrine in *Ward* v. *Rock Against Racism* (1989, 491 U.S. 781) by way of *Ward*'s quote of *Clark* v. *Community for Creative Non-Violence* (1982, 468 U.S. 288). *Grayned* was mentioned in the same paragraph as *Clark* but was not the cited source for the narrow tailoring language. The *Clark* Court never cited *Grayned*; instead, it cited a series of cases that cited *Grayned,* finally pointing to the allegedly original headwaters of the language in question. The *Clark* Court then muddied the waters even further by grounding the narrow tailoring and substantial government interest ideas in *United States* v. *O'Brien* (1968, 391 U.S. 367), which, as one could likely guess by now, the *Grayned* Court never cited!

As one further complication, the *Grayned* Court never claimed to be the original headwaters for the narrow tailoring idea; it cited twelve other cases in the two footnotes for the relevant quote, many of which are cited by other cases in the stream examined to this point. Richards and Kritzer (2002) offered a potential explanation for this complicated web of

precedents by noting that the Court in *Grayned* and in its companion case *Police Department of Chicago* v. *Mosley* (1972, 408 U.S. 92) established a content neutrality jurisprudential regime by adding a second track of analysis to the already existing content-based analysis that called for narrow tailoring to a significant government interest. In later work, Richards noted that "by examining the development of law, the network of precedents and the various justifications used by the justices, it becomes apparent that the content neutrality regime did not suddenly materialize in *Grayned* and *Mosley*" (2013, 81). In other words, the development of legal doctrine is an immensely complicated process and understanding the streams of language that ultimately shape legal doctrine requires analysis that goes beyond tracing the path of precedents through citations.

Why Study Doctrinal Language?

The development and transmission of legal doctrine is substantively interesting because doctrine forms the backbone of the law. Justices and judges often cite precedents as a way to shape the law going forward by drawing on past doctrinal developments and applying those developments to the case facts at hand. Depending on one's perspective of judicial decision-making, doctrine either governs case outcomes or instrumentally legitimizes case outcomes. In either case, the doctrine of a case is what ultimately matters for the future of the law. However, as this brief case study has demonstrated, doctrines often develop independently of explicitly acknowledged precedents. Understanding the various ways in which legal doctrine shapes the development of the law requires identifying ways to study doctrines in and of themselves. This process can then answer a variety of questions about doctrine: do doctrines have lifespans, and can they be empirically modeled? Are there attitudinal and legal variables that affect the lifespan

of a doctrine and the choice to apply a doctrine in a case? How does doctrinal choice affect the development of the law?

Since legal doctrines are expressed in the language of court opinions, answering these questions presents the challenge of measuring the use of language in empirically useful ways. This challenge has often divided scholars of the attitudinal and legal models of decision-making. Responding to criticism of *The Supreme Court and the Attitudinal Model Revisited* offered by Spriggs, Segal and Spaeth argued that the attempt by Spriggs to separate voting outcomes from policy outcomes represented "nothing more than a craving to measure a qualitative matter (words) quantitatively," adding that they "wish[ed] him well" (2003, 35). This skepticism toward studying judicial language is nothing new; as early as 1965, Spaeth differentiated himself from Mendelson by noting that he chose to focus on voting behavior rather than Mendelson's choice to focus on opinion content because votes could be trusted and words could not (Spaeth 1965). However, theoretical and empirical insights from the so-called "doctrinal politics" literature and the methodological tools of automated content analysis provide new possibilities for the challenge of modeling the creation and lifespan of legal doctrines and their impact on the development of the law (and consequently, judicial behavior). In this dissertation, I leverage a series of automated content analysis tools to explore how legal doctrine develops and is utilized by the Supreme Court and circuit courts.

In some ways, this project picks up where Hansford and Spriggs left off in their 2006 book *The Politics of Precedent on the U.S. Supreme Court*. In the book's introduction, the authors noted that their measure of the interpretation of precedent may be open to criticism because it is an imprecise operationalization of the changing meaning in precedent over time. They "trade off the ability to capture nuances in the content of law as it changes over time for the

ability to test rigorously the predictions from [their] theoretical model" (Hansford and Spriggs 2006, 15). Their work depends on evaluations of the interpretation of precedent developed by *Shepard's Citations*. In this project, I utilize the tools of automated content analysis to trace the development of doctrine through the actual text of opinions, rather than evaluations of those opinions. From a theoretical perspective, I build upon their model with an argument that the measure of the utilization of legal doctrine should be conceptually distinguished from the measure of the interpretation of precedent. From a methodological perspective, I demonstrate how automated content analysis can be used to capture more of the nuance of the content of law while rigorously testing a model of doctrinal utilization. Substantively, I measure the impact of factors such as doctrinal age, judicial ideology, and hierarchical effects upon the decision to utilize legal doctrine.

A Preview of the Dissertation

In chapter two, I identify theoretical and methodological issues that exist in previous attempts to explain precedent citation. I argue for an important distinction between the broader concept of precedent and the narrower concept of legal doctrine. I show how the reliability and validity of *Shepard's Citation*s is questionable if used as a tool to track the development of legal doctrines. This is a result of *Shepard's* dependence upon explicit citations (which is not useful in instances like the aforementioned *Grayned* case study) and a result of the coding rules that *Shepard's* adopts to classify the interpretation of precedent. These issues are highlighted through a detailed case study of the *Lemon* test in Establishment Clause jurisprudence.

In chapter three, I address these issues through the development of the concept of doctrinal vitality, a novel modification of the precedent vitality measure created by Hansford and Spriggs. As the measurement of doctrinal vitality is dependent upon the ability to measure the

linguistic composition of court opinions, I demonstrate the challenges presented by previous methods of measuring opinion language in the judicial politics literature and argue for unsupervised topic modeling and supervised sentence counts as a promising alternative.

In chapter four, I test this linguistic modeling methodology using two key doctrines identified by the jurisprudential regimes literature (content neutrality and the *Lemon* test). Using these doctrinal areas (and the statutory area of Title VII employment discrimination governed by *McDonnell Douglas*), I explore factors that affect the decision to utilize doctrine on the Supreme Court level.

In chapter five, I examine doctrinal utilization on the circuit court level. From a theoretical perspective, I explore how hierarchical effects and horizontal stare decisis should impact doctrinal utilization. To test these expectations, I construct a dataset of published and unpublished Establishment Clause circuit court opinions and use the text analysis methods developed in the Supreme Court chapter to examine utilization of the *Lemon* test. I close with a brief concluding chapter that highlights overall key takeaways from the project and suggests areas for future research.

CHAPTER TWO: THEORETICAL ISSUES WITH THE MEASUREMENT OF PRECEDENT AND DOCTRINE

In this chapter, I argue for the need to distinguish between the concepts of precedent and legal doctrine when tracking the development of the law. I begin with a brief discussion of the role that stare decisis plays in the development of the law and explore various definitions of the concept of legal doctrine. I present a theory of doctrinal development and use that theory as a vehicle for demonstrating the way in which "precedent" is an umbrella term that encompasses various concepts that must be conceptually distinguished. I argue that a precedent contains three major components: the case facts, the *ratio decidendi* (decision rule), and the case outcome. The decision rules are the legal doctrine that shape the development of the law, and thus are worthy of study in and of themselves. However, studies that analyze the interpretation of precedent collapse the components of a precedent together and thus demonstrate the posture that a court takes toward a particular combination of facts, doctrine, and outcome. This is a conceptually distinct question from the way in which legal doctrine is utilized by courts.

To demonstrate this distinction, I provide a detailed case study of the development and application of the *Lemon* test to show how previous tools for measuring the interpretation of precedent present difficulties for measuring the development of legal doctrine. Specifically, I show instances in which the Supreme Court utilized the *Lemon* test while taking a neutral or negative posture toward the precedents that established it (e.g., in instances where the case facts do not align with each other). This distinction points to the need for methods that enable the study of legal doctrine as its own constituent element of opinions.

Stare Decisis and Doctrinal Development

Judges and justices cite precedents for two primary reasons: to assist in the communication of their policy preferences and to buttress the legitimacy of their decisions (Hansford and Spriggs 2006, 12). While most scholars would accept these two basic functions of precedent, there is a divide between the two broad understandings of judicial decision-making as to the extent to which precedent possesses an independent constraining influence while serving these functions. Proponents of the attitudinal model treat precedent as purely instrumental, whereas proponents of the legal model see precedent as influencing case outcomes.

As attitudinalists, Segal and Spaeth (2002) argued that precedent is not constraining due to the existence of usable precedents on both sides of any legal question that appellate courts hear. Justices possess a preferred policy outcome driven by their personal attitudes and vote for the side that best actualizes that outcome. However, institutional features of the judiciary force justices to appear to rationalize this choice through appeal to stare decisis (Hansford and Spriggs 2006), which they can do through appeal to a menu of doctrinal options on either side of the case. Judges risk reputational costs if they fail to abide by the norm of stare decisis (Miceli and Coşgel 1994), because failure to abide by the norm negatively impacts the legitimacy of both the judge and the opinion. Opinions of questionable legitimacy are less likely to be seen as strong precedents by future courts, and judges of questionable legitimacy are less likely to be cited as authoritative. Thus, legitimacy directly impacts the ability of a judge to see personal policy preferences implemented. Attitudinalists acknowledge this normative expectation to ground one's decision in the logic of previous cases but see it as nothing more than a rhetorical game that justices play. If they simply announced a new doctrine without connecting it to precedent or previously created doctrine, their decision would lack legitimacy and would undermine the

legitimacy of the judiciary as an institution. To sum up Segal and Spaeth's view of stare decisis: "Norm? Yes; Influence? No" (1996, 1074).

Conversely, the legal model takes justices and judges seriously when they claim to be influenced by the force of doctrine as derived from precedents. Modern advocates of the legal model (e.g., Gillman 2001; Tamanaha 2009) do not dismiss judicial attitudes as irrelevant, but portray those attitudes as conditioned by legal factors. In terms of the two broad functions of precedent: adherence to the legitimizing norm of stare decisis places some constraints on the ability of justices and judges to communicate their policy preferences. Justices or judges who want to move legal doctrine closer to their preferred policy outcome while obeying the norm of stare decisis are forced to at least start from a precedential doctrine and modify it as desired. While justices may be playing a rhetorical game, the game is not just instrumental, but strategic. Justices and judges have preferred policy positions, but must strategically modify those positions to attempt to shift legal doctrine while starting from previously established precedent (Knight and Epstein 1996). Segal and Spaeth are correct in their assertion that precedents present a menu of legal doctrines, but menu options are constraining nonetheless. I contend that referencing a precedent (even if for purely instrumental reasons) inevitability impacts the development of legal doctrine in that issue area.

What is Legal Doctrine?

Legal doctrine expressed in precedents is constraining because it sets the boundaries of acceptable case outcomes. Cohen interpreted the meaning of doctrine in the positive political theory model of McNollgast (1994) by defining doctrine as a "zone of discretion within which lower court decisions will not be reversed by the Supreme Court" (1994, 1686). Tiller and Cross similarly stated that "in many respects, doctrine, or precedent is the law, at least as it comes from

courts. Judicial opinions create the rules or standards that comprise legal doctrine…Legal doctrine sets the terms for future resolution of cases in an area" (2006, 517). Doctrine is the instrument by which case outcomes are mapped onto case facts (Jacobi and Tiller 2007). Richards and Kritzer (2002) theorized a similar function for jurisprudential regimes, as regimes identify the relevant case factors and/or balancing tests that shape case outcomes. Jurisprudence, driven by the norm of stare decisis, is thus a process of reasoning by example in which a judge attempts to determine the appropriate alignment between case facts and precedents and then apply the relevant legal doctrine.

To illustrate how the development of legal doctrine is a process of determining the boundaries of the case space within which a doctrine applies in addition to delineating acceptable case outcomes, consider the classic model of legal reasoning articulated by Levi (1948). In his theory, doctrine has a three-stage lifespan: creation, stability, and breakdown. Doctrine is at its most unstable and indeterminate in the creation stage; this is when reasoning by example is used to clarify the boundaries of the case space. Wahlbeck (1997, 1998) similarly argued that the development of a legal rule involves the transfer of case facts into and out of the scope of the rule (in case space terms, a widening and narrowing of the boundaries of the space). This is a dual-development process. Reasoning by example defines both the extent of the case space in which the doctrine applies and clarifies the decision rule articulated by the doctrine. For example, in Establishment Clause jurisprudence, the development of the boundaries of the case space is the process of defining the set of cases to which the *Lemon* test applies, whereas the process of clarifying the decision rule involves defining the outcome that should be produced when the *Lemon* test is applied to a particular case.

Cardozo, referring to the modification of legal doctrine on a case-by-case basis, wrote that "this modification is gradual. It goes on inch by inch" (1921, 25). Once enough examples are accumulated to establish the boundaries of the case space and the application of the decision rule with sufficient clarity, the doctrine transitions into the stability stage. The breakdown of a doctrine occurs when a judge determines that new sets of case facts produced by legal, technological, or sociological evolution can no longer fall within the case space governed by the doctrine. Thus, doctrine gains meaning over time, and that meaning is transmitted via precedent. Bueno de Mesquita and Stephenson (2002) theorized a similar role for precedent as serving a communicative function that clarifies the meaning of a doctrine. This communicative function can happen vertically as well, as the Supreme Court examines how lower courts are interpreting a precedent to determine if the Court needs to distance itself from that precedent in future cases (Hansford, Spriggs, and Stenger 2013).

The following figures present this theory of doctrinal development and precedential communication in a stylized fashion. In each figure, the outer circle represents the case space that a legal doctrine covers. The line that divides the case space is the decision rule or rules articulated by the doctrine. The solid circles represent individual cases (or, more precisely, unique collections of case facts). The figures progressively represent the lifespan of a doctrine. Figure 1 represents the development of a new legal doctrine.

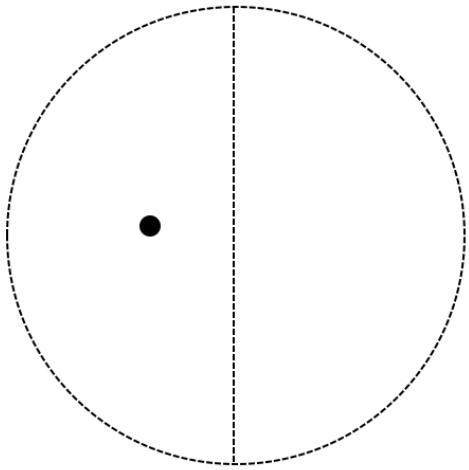


Figure 1. Unstable case space boundaries, unstable decision rule

The solid circle is the first set of case facts that a court determines is covered by the doctrine. The decision rule divides the case space into two halves. The court uses the decision rule to determine on which side of the case space the particular set of case facts falls. However, since this is the first case in which the new doctrine is being applied, both the meaning of the decision rule and the boundaries of the case space are uncertain (as indicated by the dashes).

Once a court decides that a second case is governed by a legal doctrine, the placement of the doctrine within the case space helps to clarify the meaning of the decision rule. This is part of the communicative function of precedent. The court is communicating to future courts two implications of the doctrine: the set of case facts that are covered by the doctrine, and the outcomes that are produced by applying the appropriate decision rule. At this point in the doctrinal development process, both sets of case facts fall within the scope of the doctrine, but the outcomes fall on opposite sides of the decision rule. This means that the decision rule is beginning to stabilize, as it becomes clearer how the decision rule should be used to produce case

outcomes. The decision rule line remains dashed to indicate that the rule is stabilizing but has not reached its final stability point because future cases can still clarify the rule (see figure 2).

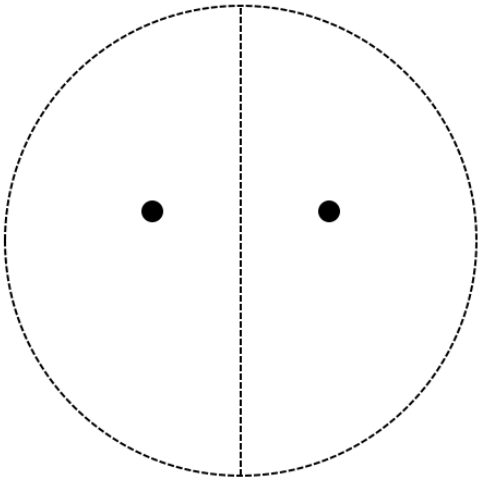


Figure 2. Unstable case space boundaries, stabilizing decision rule

The case space boundary remains unstable because the court has only defined cases that fall within the boundary; it has not yet identified a set of cases within the issue area but outside of the doctrinal boundary. In Levi's terms, the doctrine is still in its creation phase, because neither the decision rule or case space boundaries have been stabilized.

As the doctrine continues to be applied in the issue area, the meaning of the decision rule stabilizes (note that the dividing line in figure 3 is now solid).

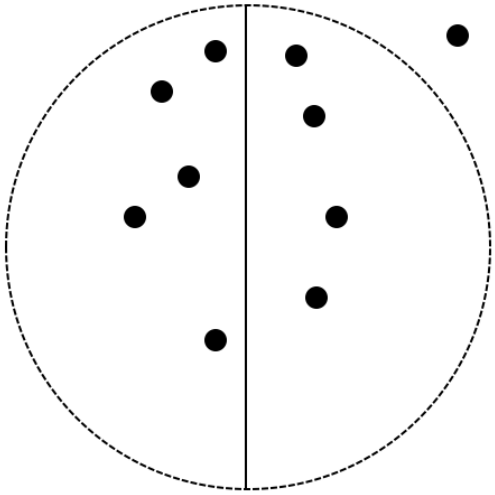


Figure 3. Stabilizing case space boundaries, stable decision rule

Several sets of case facts have been placed on both sides of the case space, which means that the courts have used precedents to clearly communicate how a decision rule should be interpreted and applied to produce certain case outcomes. However, note a new development on the outside of the case space. The court has encountered a set of case facts in the issue area that does not fall within the case space demarcated by the doctrine; thus, these case facts are placed outside of the space of the current doctrine. A new case space could be drawn around the outsider set of case facts to represent the development of a new doctrine; however, these figures focus only on the lifespan of the original doctrine. The case space boundaries remain unstable (dashed) because more sets of case facts are needed to make it clear which precedents will be governed by the new doctrine and what the demarcation line between doctrines is. As the case space boundaries begin to stabilize, the doctrine begins its transition from creation to stability.

In the final stage of the doctrine's development, both the decision rule and the case space boundaries have been stabilized (as indicated by the solid lines in figure 4).

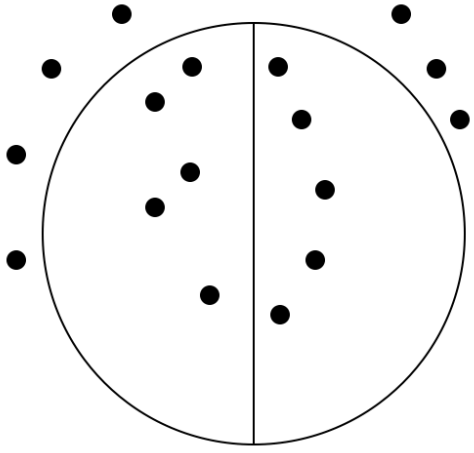


Figure 4. Stable case space boundaries,
stable decision rule

The court has now clearly communicated which sets of case facts fall within and outside the governance of the doctrine. New doctrines have been developed to govern case facts that fall outside the case space of the original doctrine. The court could decide to group all of the new

cases under a single new doctrine or develop multiple doctrines to cover the cases. In terms of Levi's model of legal development, the doctrine is now transitioning from stability to decline. The court has decided that the doctrine is no longer useful for new sets of case facts (that have come about as a result of legal, technological, or social evolution, as noted earlier). Throughout this developmental process, it is clear that precedents are being used to communicate both the meaning of the decision rule and the boundaries of the case space.

Segal and Spaeth critiqued this communicative function of precedent on two grounds: the malleability of linguistic meaning and the malleability of the zone of discretion set by precedent. As part of their criticism of the "plain meaning" component of the legal model, they argued that understanding the plain meaning of a statute or judicial opinion is impossible because language "possesses a chameleonic quality that spans the color spectrum. First, English as a language lacks precision. Virtually all words have a multiplicity of meanings, as the most nodding acquaintance with a dictionary will attest" (Segal and Spaeth 2002, 54). In other words, the decision rule is never stabilized. However, Bueno de Mesquita and Stephenson (2002) theorized that justices cite chains of precedent to clearly communicate how they interpret a doctrine. Justices certainly engage in the interpretation of precedent, but since precedents possess a limited range of meaning, there are certain interpretive boundaries that justices cannot cross, lest they delegitimize institutional norms such as stare decisis. To modify Segal and Spaeth's metaphor, words may have a chameleon-like quality, but they can only change to a limited amount of colors (and those colors become more narrowly defined over time as the zone of discretion gains clarity through intrajudicial communication).

Segal and Spaeth's second critique of the constraining power of precedent concerned the flexible boundaries of a doctrine. They argued that justices are free to choose from a selection of

precedents that speak to the facts of a case. Since doctrines map outcomes onto facts, judges and justices can strategically manipulate the facts that they apply as relevant in a case to produce the outcome that they desire, a choice that is again driven by attitudes (Segal and Spaeth 2002, 80). The boundaries of the case space never stabilize, because justices can consistently redefine the cases that are covered by a doctrine. However, this is precisely the point for advocates of the legal model: even if doctrine is being used instrumentally as an expression of attitudes, a doctrine must still be chosen. The doctrinal instrument that is used matters. Or, as Schauer put it, "constitutional doctrine and constitutional precedent channel behavior and thinking in the direction established by that doctrine and that precedent" (1984, 664). If doctrine *b* is chosen rather than doctrine *a*, this will modify the outcome of the case in terms of the law that it promulgates in some manner (even if that manner is ever-so-slight and is driven primarily by judicial attitudes). The two doctrines might emphasize different dimensions of the case facts or may establish different rules and standards to follow. The change might not be enough to change the outcome of the vote on the merits, but doctrine is ultimately promulgated through opinions, not the vote on the merits (Bartels and O'Geen 2015), and the outcome of the vote on the merits can only tell us so much about the ideological content of an opinion (McGuire et al. 2009).

For this reason, judges and justices care about the substantive doctrinal content of opinions. Disagreement over the framing of controlling statutes or the issue area of a case has been linked to the decision to dissent rather than concur (Benesh and Spaeth 2007). Smith and Tiller (2002) found that judges strategically specified the "appropriate" legal doctrine in administrative law cases to increase the likelihood of their preferred policy outcome. Smith (2014) argued that lower court judges strategically take advantage of the lower likelihood of fact-bound cases being reviewed when compared to law-bound cases by writing opinions in fact-

centric manners, thus manipulating which doctrinal standard of review is more likely to be utilized. Subdividing doctrine into the more concrete "rules" and more ambiguous "standards," Smith and Todd (2015) found that rules provide greater constraint on lower court judges. Thus, the study of doctrine as a strategic instrument of judicial policy is important from two perspectives: its influence upon case outcomes and the choice of which doctrines to utilize.

The rules/standards distinction gives explicit meaning to the general category of legal doctrine. Rules are "strict requirements that define the answer to a dispute," whereas standards are "amorphous guides to resolving disputes, often listing a set of factors to be considered and balanced" (Tiller and Cross 2006, 517). Rules narrow the zone of discretion that a judge possesses in interpreting a doctrine, whereas standards make the zone broader. Jacobi and Tiller (2007) provided the formal model that lies behind the findings of Smith and Todd (2015). A judge or justice who wants to exert greater control over lower courts should propagate a rule rather than a standard (a situation that arises, for instance, when lower courts are ideologically divergent from the higher court). Since rules are more determinate than standards, rules provide heightened opportunity for judicial oversight (Lax 2012). The study of the strategic use of doctrinal instruments to ensure the implementation of preferred policy outcomes on varying levels of the judicial hierarchy has been labeled the "doctrinal politics" approach. Advocates of the approach make the important claim that "judicial preferences and goals are structured by the choices available to judges: dispositions and the rules that denote general sets of case dispositions" (Lax 2011, 133). Even Segal himself has softened his stance on the importance of legal doctrine. D'Elia-Kueper and Segal (2015) argued that judges can use legal doctrine as an instrument to promote their preferred outcomes, a reality that underlies the precedent vitality model of Hansford and Spriggs (2006).

Distinguishing Doctrine and Precedent

Legal doctrine maps case outcomes onto case facts. A precedent thus consists of three major parts: the case facts, the legal doctrine through which the case facts are evaluated, and the case outcome produced because of that evaluation. When building their theory of precedent vitality, Hansford and Spriggs collapsed these concepts when they defined precedent as the "legal doctrines, principles, or rules established by prior court opinions" (2006, 5). I argue that this definition fails to note that a precedent also contains a set of case facts, and this omission explains why their model is underinclusive as a measure of legal doctrine. Of course, Hansford and Spriggs are not alone in conflating precedent and doctrine; Tiller and Cross (2006) also defined the concepts synonymously. I contend that a precedent is not a doctrine in and of itself; rather, a precedent is the *application* of a doctrine. Thus, the vitality of a precedent and the vitality of a doctrine are two separate matters. Precedent vitality captures the extent to which a particular combination of case facts and legal doctrines influence future decisions. To portray this distinction visually, I have reproduced the original case space with a single precedent but including labels to distinguish between the components of a precedent (see fig. 5).

As noted, the case space is the range of cases in which a particular legal doctrine is applicable. Within that case space, each case presents a certain set of case facts that are evaluated according to a decision rule or rules articulated by the legal doctrine. The rule or rules divides the case space into two possible outcomes (for instance, pro-plaintiff or pro-defendant). A precedent presents a certain set of case facts that exist within the boundaries of a case space in which a

legal doctrine applies, and the case facts are distilled through decision rules to produce the possible outcomes within the case space.

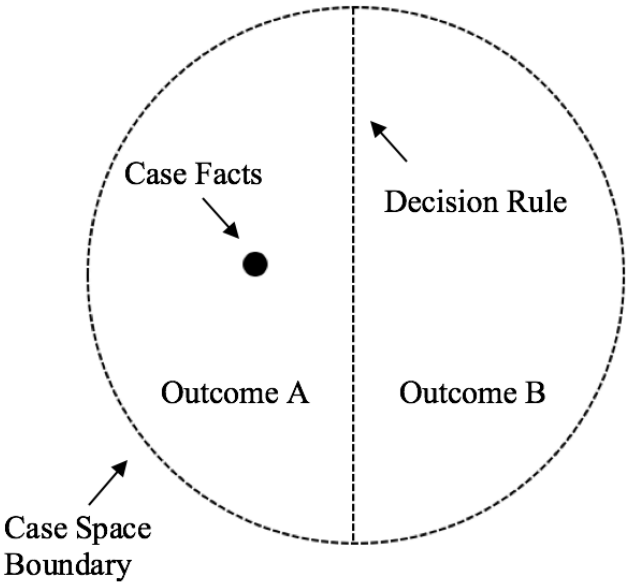


Figure 5. The components of a precedent within a case space

Thus, it is inappropriate to think of a case outcome as being controlled by a particular precedent. The case outcome is actually controlled by a decision rule. However, since the goal of legal reasoning under the norm of stare decisis is to identify a precedential set of case facts that align with the case at hand and use that case's application of the decision rule to select the appropriate case outcome, cases often *appear* to be controlled by a particular precedent. In some instances, the court chooses to overrule a prior precedent, which can occur when the court determines that the prior court applied the correct decision rule to the given case facts but misapplied the rule and thus produced the incorrect outcome. This can also occur when the court decides that the prior decision rule was wrong.

An Example of Doctrinal Development – The *Lemon* Test

When the case facts and decision rule components of a precedent are conceptually distinguished, it becomes apparent that another route of weakening a precedent is possible. To see this route in action, consider the case space in figure 6, which contains four foundational Establishment Clause cases from the Supreme Court.

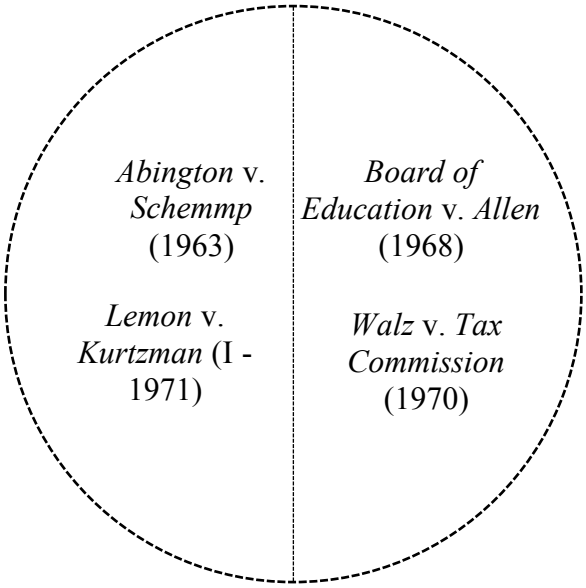


Figure 6. The *Lemon* test case space as of *Lemon* I

The legal doctrine represented by this case space is the three-pronged *Lemon* test. Under this test, a governmental action does not violate the Establishment Clause of the First Amendment if it has a secular legislative purpose, does not have the primary effect of advancing or inhibiting religion, and does not promote excessive government entanglement with religion. The test is referred to as the *Lemon* test because the three prongs were combined into one test in *Lemon* v. *Kurtzman* I (403 U.S. 602), decided in 1971 (hereafter, referred to simply as *Lemon* although there was a follow-up *Lemon* v. *Kurtzman* two years later).

The evolution of the *Lemon* test demonstrates the complicated relationship between precedent and doctrine. The first two prongs (the purpose and effect prongs) of the *Lemon* test

were originally articulated in the 1963 case *Abington* v. *Schempp* (374 U.S. 203). In *Abington*, the Court ruled that a Pennsylvania law that required the reading of the Bible at the beginning of each school day was an unconstitutional establishment of religion because Bible reading was a religious exercise and therefore possessed a religious purpose that was not neutral toward religion. The purpose and effect prongs were applied again in the 1968 case *Board of Education* v. *Allen* (392 U.S. 236). In *Allen*, the Court ruled that a textbook distribution program in which the state allowed students who attended religious schools to receive free textbooks did not violate the Establishment Clause because the program had the secular purpose of promoting education Additionally, the Court ruled that the program did not advance religion because only secular textbooks were provided. The Court determined that the benefits of the program flowed to the students and parents, not to the religious schools. If one was to draw an Establishment Clause case space at this point in the development of the doctrine, the decision rule would be the purpose and effect prongs, with *Abington* on the "liberal" (pro-separationist) and *Allen* on the "conservative" (pro-accommodationist) side.

The *Allen* Court pointed to the communicative function of precedent as a tool for clarifying the interpretation of doctrine when it noted the *Abington* Court's use of the example of *Everson* v. *Board of Education* (1947, 330 U.S. 1) as an illustration for understanding the proper use of the purpose and effect prongs. According to the *Allen* Court, by citing *Everson*, the *Abington* Court demonstrated a program that passed constitutional muster under the general principles of the two prongs, even if the two prongs had not yet been articulated. Using that understanding of the purpose and effect prongs, the *Allen* Court placed the case facts on the opposite side of the case space from *Abington*. In *Abington*, the law promoted religion, so it was unconstitutional; in *Allen*, the program was purely secular, so it passed muster. The same

decision rule was used, but the case facts were different, thus producing a different outcome. However, it is the way in which the case facts were different that is important for my theory of distinguishing between precedent and doctrine. The facts in *Allen* were not in direct contrast with the facts in *Abington*; *Allen* addressed a financial support issue, whereas *Abington* addressed a religious activity. Thus, while *Allen* used the same decision rule as *Abington*, it did not depend upon *Abington* as a controlling precedent but existed in the same doctrinal space.

Two years after *Allen*, the Court further complicated the doctrinal space in *Walz* v. *Tax Commission* (1970, 397 U.S. 664). Writing for the majority, Chief Justice Burger used somewhat of a hybrid of the purpose and effect prongs and created a modified effect prong known as the entanglement prong. Burger wrote that the "legislative purpose of the property tax exemption is neither the advancement nor the inhibition of religion" (*Walz* v. *Tax Commission* 1970, 672). This analysis appears to conflate the requirement that a statute have a secular legislative purpose with the requirement that the statute not have the primary effect of advancing or inhibiting religion (although Burger does cite the purpose and effect prongs when discussing *Allen*). Burger then proceeded to articulate a different requirement for the effect of the statute: "We must also be sure that the end result -- the effect -- is not an excessive government entanglement with religion" (*Walz* v. *Tax Commission* 1970, 704). Rather than using the effect test to evaluate the advancement or inhibition of religion, Burger used it to evaluate the possibility of excessive entanglement with religion. Using those tests, the Court determined that property tax exemptions for churches were constitutional, because the state was not transferring any funding to the churches. Again, the fact pattern differed in several important ways from either *Abington* or *Allen*, but the Court found a way to modify the decision rule while placing *Walz* within the *Abington/Allen* case space.

The various prongs of the decision rule were brought together in *Lemon* v. *Kurtzman*, decided the year after *Walz*. Writing for the majority, Chief Justice Burger combined the purpose and effect prongs from *Allen* and the entanglement prong from *Walz* to create the three-prong *Lemon* test. Applying two of the prongs of the test, the Court determined that direct financial support to religious schools (in the form of textbook funding, instructional materials, and teacher salaries) was an unconstitutional establishment of religion. While the programs had a clear secular legislative purpose, the attempt to separate secular and religious educational functions in the programs created an excessive government entanglement with the administration of religious schools. Since the programs failed the entanglement prong, it was unnecessary to evaluate the programs against the effect prong. While *Lemon* was controlled by the excessive entanglement prong from *Walz*, the Court distinguished the two fact patterns because of differing amounts of historical evidence as to the impact of church property tax exemptions versus the impact of government funding for religious schools. The Court also distinguished *Lemon* from the fact pattern of *Allen* because of the distinction between financial support going directly to schools versus to students and parents. However, although the fact patterns were distinguished, the programs in *Allen* and *Lemon* both passed the purpose prong.

In short, the etiology of the *Lemon* test is complex. Two prongs were initially articulated in *Abington* and re-articulated in *Allen*. *Walz* used a modification of those two prongs and added an entanglement test to the effect prong (while citing *Allen*, not *Abington*, as the source of the purpose and effect prongs). *Lemon* reverted to the purpose and effect prongs as articulated in *Allen* and separated excessive entanglement from the effect test of *Walz* to make excessive entanglement its own third prong. In my case space model terms, *Lemon* combined the decision rules of *Allen* and *Walz* while distinguishing itself from the case facts of the two precedents and

articulating an outcome on the opposite side of the case space. Thus, while the case facts of *Allen* and *Walz* were narrowed in terms of their controlling applicability, the decision rules were expanded to have controlling applicability to a broader range of case facts. Previous measures of precedent interpretation focus upon the zone in which the case facts of a precedent apply while potentially underrepresenting the zone in which the decision rule applies.

Shepard's Citations and the Problem of Precedent Vitality

Hansford and Spriggs (2006) developed the standard measure of precedent interpretation in the literature. This measure, known as precedent vitality, is a running sum of the positive and negative interpretations of a precedent by future opinions. To code positive and negative interpretations, Hansford and Spriggs depended upon the coding of *Shephard's Citations*. In addition to evaluating the Supreme Court's use of its own precedent, *Shepard's* has been used to evaluate compliance with Supreme Court precedent by state supreme courts (Fix, Kingsland, and Montgomery 2017; Kassow, Songer, and Fix 2012) and federal circuit courts (Hinkle 2017). Under this schema, a positive interpretation of a precedent occurs when a case "follows" a previous case; the "follow" designation is only applied when a case cites a previous case as a "controlling authority" or "compelling precedent" (Hansford and Spriggs 2006, 45, quoting the in-house *Shepard's* training manual). *Shepard's* designations are based on linguistic signals provided in opinions.

For the Court to signal that a precedent is a controlling authority, two conditions must be met. First, the precedent must be explicitly cited. Precedents that are either tangentially related to the core issue of a case or contain language that may have influenced the opinion without being cited by the opinion are not coded as vital precedents. Hansford and Spriggs admitted that this dependence on "explicit" treatment of a precedent may be "somewhat underinclusive" (2006,

50). Secondly, if a precedent is explicitly cited, the Court must determine that the case facts and the legal doctrine of the citing case align with the precedential decision. If a precedential doctrine is applied, but in a case where the fact pattern differs from the precedent, then the Court may apply the *doctrine* without following the precedent as a controlling authority.

The development of the *Lemon* test provides a great example of both of these issues. Recall that although *Lemon* quotes the purpose and effect prongs from *Allen*, the prongs were originally articulated in *Abington*. The only citation to *Abington* in the *Lemon* majority opinion is a citation to Justice Goldberg's concurring opinion. Thus, citation network analysis (Fowler and Jeon 2008; Fowler et al. 2007) does not present a complete picture of how doctrine develops. There are other examples in Establishment Clause jurisprudence where citation practices cause the influence of *Lemon* to be missed by the precedent vitality measure. In *Rosenberger* v. *Rector and Visitors of University of Virginia* (1995, 515 U.S. 819), while not citing *Lemon*, the Court quoted the concept of entanglement from *Widmar* v. *Vincent* (1981, 454 U.S. 263), which "follows" *Lemon*.[1] In *Harris* v. *McRae* (1980, 448 U.S. 297), the Court quoted the *Lemon* test, but from *Committee for Public Education and Religious Liberty* v. *Regan* (1980, 444 U.S. 646), and never cited *Lemon*.

Precedent vitality also mispresents the influence of *Allen* and *Walz* upon the *Lemon* test. *Shepard's* codes *Allen* as being "distinguished" by *Lemon*, due to this excerpt from *Lemon*: "This factor distinguishes both *Everson* and *Allen*, for in both those cases the Court was careful to point out that state aid was provided to the student and his parents -- not to the church-related

[1] To further complicate matters, *Rosenberger's* quote from *Widmar* regarding entanglement cites *Walz*, (the original source of the entanglement prong), not *Lemon*! (see *Rosenberger* v. *Rectors of the University of Virginia* 1995, 845).

school" (*Lemon* v. *Kurtzman* 1971, 621).[2] Since *Allen* possesses this key fact pattern difference, the Court distinguished *Allen* as not directly applicable to the facts of *Lemon*, but utilized one of the prongs from *Allen* to develop the doctrine that is applicable to the facts of *Lemon*. *Shepard's* treats "distinguished" as a negative treatment category (although weakly so – see Hansford and Spriggs 2006, 50); consequently, the only cite in *Lemon* that explains two of three prongs of the *Lemon* test led to its source receiving a negative vitality score! As noted earlier, the third prong of the test is derived from *Walz*, which is also distinguished from the case facts of *Lemon*; thus, the entirety of the doctrine is derived from precedents "negatively" interpreted by *Lemon*.

Additionally, since precedent vitality is concerned with controlling authorities, neutral cites are ignored. This decision can make tracking doctrine more difficult. In *United States* v. *Scott* (1999 U.S. App. LEXIS 39132), the Fifth Circuit followed *Lemon* while arguing that Justice O'Connor had subsumed the entanglement prong under the effect prong in the 1997 case *Agostini* v. *Felton*. This reference was enough for *Shepard's* to code the Fifth Circuit as "abrogating in part as stated in," meaning that *Lemon* "has been effectively, but not explicitly, overruled or departed from by an earlier decision."[3] Kritzer and Richards (2003) noted that Justice Thomas called *Agostini* a "revision" of the effect prong in his plurality opinion in *Mitchell* v. *Helms* (2000, 530 U.S. 793). The coding scheme of *Shepard's* treats *Agostini* as neutrally citing *Lemon*; thus, dependence upon precedent vitality ignores the significance of *Lemon* for the doctrinal revision that occurs in *Agostini*. To be fair, Hansford and Spriggs have demonstrated the validity of *Shepard's Citations* as a method for capturing the interpretation of

[2] Specifically, *Shepard's* defines the "distinguished" coding as appropriate in the case of "dissimilar facts" or a "different application of the law."
[3] However, *Lemon* also received a "follow" code, as other aspects of *Lemon's* logic were used in *Scott*.

precedent (Hansford and Spriggs 2006; Spriggs and Hansford 2000), but I argue it is inadequate as a measure of the conceptually distinct concept of doctrine. In case space terms, precedent vitality captures the extent to which two sets of case facts are proximate to one another; it does not capture the extent of the area that a legal doctrine covers.

Establishment Clause Case Study – *Shepard's Citations*

According to *Shepard's*, fifty-five majority opinions cite *Lemon*. Six of those opinions cite *Lemon* to distinguish the current case from *Lemon*; however, four of those six cases utilize the *Lemon* test as the appropriate decision rule. In *Tilton* v. *Richardson* (1971, 403 U.S. 672), an opinion announced the same day as *Lemon* and also written by Chief Justice Burger, the Court held that construction grants to church-related colleges for constructing a non-religious building were constitutional, because the program passed all three prongs of the *Lemon* test (as long as a twenty-year prohibition on religious use for buildings was converted to a permanent prohibition). However, since the programs in *Lemon* required greater government surveillance than the grants in *Tilton*, *Tilton* was distinguished from *Lemon*, even as the same decision rule was used. A similar distinguishing of fact pattern occurred five years later in the 1976 case *Roemer* v. *Board of Public Works of Maryland* (426 U.S. 736). While using the *Lemon* test to evaluate a Maryland program that provided grants to colleges, including religious colleges, the Court determined that the program was constitutional because the details of the institutional use of the grant were much closer to *Tilton* than they were to *Lemon.* Thus, *Shepard's* codes *Roemer* as "following" *Tilton* while being distinguished from *Lemon*, even though both cases use the decision rule from *Lemon*.[4]

[4] Interestingly, Chief Justice Burger wrote both *Lemon* and *Tilton*, and voted with the majority in *Roemer*, even though *Tilton* and *Roemer* were conservative outcomes and *Lemon* was a liberal outcome. Justice Blackmun, who wrote *Roemer*, demonstrated the same voting pattern.

In the 1977 case *New York* v. *Cathedral Academy* (434 U.S. 125), the Court invalidated reimbursement payments to private schools for expenses required by state law. The Court noted that the law that authorized the reimbursements was "unconstitutional because it will of necessity either have the primary effect of aiding religion, see *Levitt* v. *Committee for Public Education*, supra, or will result in excessive state involvement in religious affairs. See *Lemon I, 403 U.S. 602*" (*New York* v. *Cathedral Academy* 1977, 133). Even while expressing this dependence on the decision rule of *Lemon*, the Court distinguished the fact patterns (the issue in *Lemon* was administration of funds, whereas the issue in *Cathedral Academy* was the funds themselves). Lastly, *Lemon* was distinguished from the 1980 case *Stone* v. *Graham* (449 U.S. 39) because *Stone* involved the posting of the Ten Commandments in public school classrooms, which had a clearly religious purpose, unlike the school funding in *Lemon*, which had the secular purpose of education. The posting of the Ten Commandments was invalidated on the grounds of the secular purpose test from *Lemon*.

In all four cases, the Court distinguished the *Lemon* fact pattern from the current case, which had the effect of limiting the relevance of a particular fact pattern but expanding the case space in which the *Lemon* test applies. The other two "distinguished" cases did limit the area in which the *Lemon* test applies, but with no clear explanation as to why. In *Cutter* v. *Wilkinson* (2005, 544 U.S. 709), the Court upheld the institutionalized persons provisions of the Religious Land Use and Institutionalized Persons Act, but explicitly stated that it was resolving the case on grounds besides *Lemon*, even though the lower courts had used *Lemon*. In *Van Orden* v. *Perry* (2005, 545 U.S. 677), the Court noted that the *Lemon* test had an uneven history of application in Establishment Clause cases, proceeded to dismiss the test as not useful for evaluating the "passive" existence of a Ten Commandments monument on the grounds of the Texas State

Capitol, and opted to evaluate the monument in light of its history and nature.[5] The case space of the six distinguished cases is illustrated in figure 7.

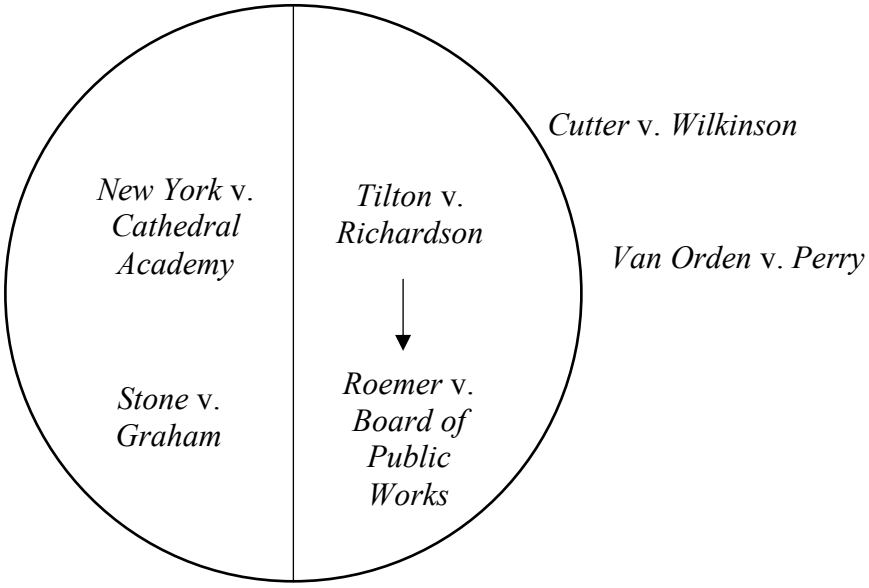


Figure 7. The relationship of the "distinguished" cases to the *Lemon* case space

The "distinguishing" represented by *Cutter* and *Van Orden* is fundamentally different than the distinguishing that occurred in the cases within the case space. Within the case space, the cases were distinguished on the basis of the case facts; outside the case space, the cases were distinguished on the basis of the extent of the decision rule. Notice that the line representing the boundary of the case space is solid, because the Court has now provided a "clear" picture of the extent of the *Lemon* test case space.[6] I have added a feature to this visualization of the case space. The arrow drawn from *Tilton* to *Roemer* indicates that *Roemer* "followed" *Tilton*, according to the coding of *Shepard's*. While the *Tilton/Roemer* relationship is a situation where

[5] On the same day in 2005, the Court issued *McCreary County* v. *ACLU of Kentucky* (545 U.S. 844), in which the display of Ten Commandments in courthouses was struck down as a violation of the purpose prong of the *Lemon* test. By the Court's logic, the test was applicable in *McCreary* because the law required the "active" placement of Ten Commandments posters.
[6] I place "clear" in quotation marks due to the ambiguity of *McCreary* relative to *Van Orden*.

the Court found a close alignment between case facts, this alignment is not strictly necessary to receive a follow designation. To demonstrate the implications of a "follow" coding, table 1 provides a summary of the twenty-one majority opinions that follow *Lemon*.

Table 1. Cases that "Follow" *Lemon*

Case Name	**Fact Pattern**	**Decision Direction**
Hunt v. *McNair* (I)	Bonds for religious colleges	Remanded in light of *Lemon*
Committee for Public Education v. *Nyquist*	Grants to religious schools	Liberal (pro-separationist)
NLRB v. *Catholic Bishop*	Jurisdiction over church schools	Conservative (not establishment)
Widmar v. *Vincent*	University facility access	Liberal
Larson v. *Valente*	Charitable reporting requirements	Liberal
Larkin v. *Grendel's Den*	Church veto over liquor license	Liberal
Mueller v. *Allen*	Education tax deductions	Conservative (pro-accommodationist)
Alamo Foundation v. *Secretary of Labor*	Fair Labor Standards Act jurisdiction	Liberal (not establishment)
Estate of Thornton v. *Caldor*	Sabbath work exemption laws	Liberal
Grand Rapids v. *Ball*	Classes for religious schools	Liberal
Witters v. *Washington*	Funding to a ministry student	Conservative
Edwards v. *Aguillard*	Act requiring creationism	Liberal
Latter Day Saints v. *Amos*	Labor nondiscrimination	Conservative
Bowen v. *Kendrick*	Grants to religious groups	Conservative
Hernandez v. *Commissioner*	Income tax deductions	Liberal
Allegheny v. *ACLU*	Religious displays	Conservative
Jimmy Swaggart Ministries v. *Board of Equalization*	Sales tax on religious materials	Conservative
Board of Education v. *Mergens*	School facility access	Conservative

Case Name	Fact Pattern	Decision Direction
Lee v. *Weisman*	Graduation prayers	Liberal
Board of Education v. *Grumet*	School district lines	Liberal
Santa Fe v. *Doe*	Graduation prayers	Liberal

By *Shepard's* coding, the Court followed *Lemon* across a variety of fact patterns. While some of the fact patterns were at least tangentially related to financial support of schools, others were not (such as the two graduation prayer cases). This analysis also points to some oddities in how the "follow" designation is applied. The Court followed *Lemon* in *County of Allegheny* v. *ACLU, Greater Pittsburgh Chapter* (1989, 492 U.S. 573), a case that related to a Chabad display and a menorah in state buildings. However, the Court distinguished *Lemon* in the aforementioned Ten Commandments case *Stone* v. *Graham*, even though the *Lemon* test was applied in both cases. Furthermore, as noted earlier, *Lemon* was distinguished in *Van Orden*, and the secular purpose prong was dispositive in *McCreary*, but *Lemon* was not followed in either case although they both dealt with religious displays! In *Lee* v. *Weisman* (1992, 505 U.S. 577), the Court followed *Lemon* without evaluating graduation prayers through the lens of the three-pronged test. The Court declined to reconsider *Lemon* but then grounded the analysis of the central issues of the case in principles distilled from *Lynch* v. *Donnelly* (1984, 465 U.S. 668) and *Everson* by way of *Allegheny* while (debatably) creating a new "coercion" test. This move led Justice Scalia to write in dissent a year later:

> As to the Court's invocation of the *Lemon* test: Like some ghoul in a late-night horror movie that repeatedly sits up in its grave and shuffles abroad, after being repeatedly killed and buried, *Lemon* stalks our Establishment Clause jurisprudence once again, frightening the little children and school attorneys of Center Moriches Union Free School District. Its most recent burial, only last Term, was, to be sure, not fully six feet under: Our decision in *Lee* v. *Weisman* conspicuously avoided using the supposed "test" but also declined the invitation to repudiate it. Over the years, however, no fewer than five of the currently sitting Justices have, in their own opinions, personally driven pencils through the creature's heart (the author of today's opinion repeatedly), and a sixth has joined

> an opinion doing so. (*Lamb's Chapel* v. *Center Moriches Union Free School District* 1993, 398, citations omitted)

Legal scholars were divided on the fate of the *Lemon* test post-*Lee*, and on the question of if coercion replaced the three prongs of *Lemon*. In an article poignantly titled "*Lemon* Is Dead," Michael Paulsen contended that the Court "has indeed interred the *Lemon* test and replaced it with a coercion test, albeit one of uncertain parameters and an uncertain future" (1993, 797). Given that the Court explicitly applied the three prongs of *Lemon* in *Lamb's Chapel* the year after *Lee*, Paulsen seemed to have prematurely pulled the plug on the test.[7] Daniel Conkle argued as much in "*Lemon* Lives", noting that the prongs of the *Lemon* test were meant to be applied in a "context-specific, case-by-case analysis of particular problems" (1993, 867).

The *Lamb's Chapel/Lee* debate is not the only time that a justice has muddied the waters relative to the use of *Lemon* in previous cases. As noted in Table 1, *Shepard's* codes *Larson* (456 U.S. 228) as following *Lemon*.[8] In *Larson*, the Court held that a rule that applied registration and reporting requirements to churches who received more than fifty percent of their contributions from nonmembers was unconstitutional because the rule was not "closely fitted" to compelling government interests and because the rule violated the excessive entanglement prong of the *Lemon* test. However, while the Court used that prong, they also noted that "although application of the *Lemon* tests is not necessary to the disposition of the case before us, those tests do reflect the same concerns that warranted the application of strict scrutiny" (*Larson* v. *Valente* 1982,

[7] While the Court invoked the *Lemon* test in *Lamb's Chapel* to evaluate the hypothetical possibility of an Establishment Clause violation, *Shepard's* does not code *Lamb's Chapel* as following *Lemon*.

[8] It should be noted this "follow" coding may be a mistake. *Shepard's* claims that the coding is based upon language found on p. 235 of the opinion; however, this language appears to refer to a lower court judge following *Lemon*, rather than the Supreme Court. Since the Court does apply the entanglement prong in its reasoning, I assume the follow code is appropriate.

252). In the same paragraph, the Court appeared to clearly distinguish *Larson* from *Lemon*; *Larson* implicated a rule that discriminated between religions, whereas *Lemon* implicated a rule that was equally applicable to all religions.[9] These two sentences perhaps influenced Chief Justice Burger when he wrote in *Lynch* that *Larson* was one of two Establishment cases decided since *Lemon* in which the Court did not "find *Lemon* useful" (*Lynch* v. *Donnelly* 1984, 679).[10] In the 1989 *Hernandez* (490 U.S. 680) case that followed *Lemon,* the Court argued that the *Larson* analytical framework was a two-step process, in which *Lemon* was applied as the second step only if the rule in question did not facially discriminate between religions. The debate over the use of *Lemon* in *Larson* is another example of the questionable application of the follow coding by *Shepard's*.

Thus far, I have identified issues with the distinguished coding and the follow coding as tools for identifying the utilization of legal doctrine. Two other coding options merit discussion: "explaining" and neutral citations. "Explaining" is the *Shepard's* code that is used when a case "interprets or clarifies" a precedent "in a significant way." Four Establishment Clause cases explain *Lemon*, but as with the other codes, a close reading of these cases reveals some specification issues. The first case to explain *Lemon* is *Hunt* v. *McNair* II (1973, 413 U.S. 734).[11] While the *Shepard's* page identification is unclear as to what language explained *Lemon*, the following sentence is a possible flag, but a curious one: "With full recognition that these are no more than helpful signposts, we consider the present statute and the proposed transaction in

[9] *Shepard's* does not code this language as distinguishing.
[10] In dissent, Justice Brennan argued that *Marsh* v. *Chambers* (463 U.S. 783) was the only case that deviated from using *Lemon*, although he also argued that a separate standard of review was adopted in *Larson* (*Lynch* v. *Donnelly* 1984, 704n11).
[11] The first *Hunt* v. *McNair* (403 U.S. 945) followed *Lemon*, as it was remanded in light of the decision in *Lemon*. This case was an evaluation of that remand.

terms of the three 'tests': purpose, effect, and entanglement" (*Hunt* v. *McNair* 1973, 741). The Court made it clear that it was evaluating the case using the *Lemon* test, but since the strength of the test was caveated with the "no more than helpful signposts" language, *Lemon* was not followed, but explained. The language implied that the test was not dispositive. A similar linguistic issue appeared in *Meek* v. *Pettinger* (1975, 421 U.S. 349), where the *Lemon* test was acknowledged as the "proper framework of analysis for the issues presented" but was also described as "guidelines" that "must not be viewed as setting the precise limits to the necessary constitutional inquiry" (*Meek* v. *Pettinger* 1975, 359).[12]

Lynch is the third case that only explained *Lemon*, and I have already identified some issues with how *Lynch* characterized the use of *Lemon* in *Larson*. *Lynch* utilized the secular purpose prong to evaluate a religious display, while making it clear that the Court did not want to be confined to a single test. Ironically, the author of the *Lynch* opinion was Chief Justice Burger, who also authored the *Lemon* test. As time passed, Burger became increasingly reticent to apply the *Lemon* test as a singularly controlling criteria, going so far as to dissent in the 1985 case *Aguilar* v. *Felton* (473 U.S. 402) by claiming that the Court's "obsession" with the *Lemon* criteria was "damaging" (1985, 419). As with *Hunt* and *Meek*, the doctrine was clearly used in *Lynch*, even if Chief Justice Burger expressed enough reticence to earn an "explained" coding. The final case that purely explained *Lemon* is *Wallace* v. *Jaffree* (1985, 472 U.S. 38), which applied the secular purpose test but further defined the test using a new endorsement test developed by Justice O'Connor. Thus, in three of the four cases that earned an explained coding,

[12] Interestingly, *Grand Rapids* v. *Ball* (473 U.S. 373) follows both *Meek* and *Lemon*, even though *Meek* does not follow *Lemon*.

the *Lemon* test was applied, and the fourth case modified the *Lemon* test while remaining within the *Lemon* framework (rather than establishing an entirely new test).

A neutral citation code is applied when a case simply cites a precedent, without any language that suggests positive or negative treatment of the precedent. As with the other categories I have analyzed, neutral citations understate the influence that a legal doctrine has upon a case. I identified several cases in which *Shepard's* codes a case as neutrally citing *Lemon*, but the *Lemon* test is applied in the logic of the opinion. To demonstrate this phenomenon, it is useful to analyze quotes from selected cases. In *Levitt* v. *Committee for Public Education & Religious Liberty* (1973, 413 U.S. 472), the Court ruled that the state could not provide funding for private school functions that were mandated by state law, because it was not possible to separate the secular and sacred aspects of private school services. The Court noted that "the essential inquiry in each case… is whether the challenged state aid has the primary purpose or effect of advancing religion or religious education or whether it leads to excessive entanglement by the State in the affairs of the religious institution." (*Levitt* v. *Committee for Public Education & Religious Liberty* 1973, 481-82). Chief Justice Burger, who wrote the opinion of the Court, linked this idea to *Nyquist* (which follows *Lemon)* and *Lemon*.

In *Sloan* v. *Lemon* (1973, 413 U.S. 825), the Court evaluated a school aid law that Pennsylvania passed in response to *Lemon* overturning the previous aid law. The *Sloan* Court ruled that the new law had the "impermissible effect of advancing religion" and that the law had no "constitutionally significant distinctions" from the law evaluated in *Nyquist*, which was decided the same day (*Sloan* v. *Lemon* 1973, 830). In *Wolman* v. *Walters* (1977, 433 U.S. 229), the Court found that a statute that enabled the state to provide various services to nonpublic school pupils was unconstitutional and noted that "the mode of analysis for Establishment Clause

questions is defined by the three-part test that has emerged from the Court's decisions" (1977, 235). *Roemer*, *Nyquist*, and *Lemon* were credited for the test.

Conclusion

This detailed case study of how *Lemon* v. *Kurtzman* has been treated by *Shepard's* reveals the flaws of using *Shepard's* to track the application of legal doctrine. *Shepard's* coding is designed to analyze the interpretation of a precedent, rather than the legal doctrine that is one of the components of the precedent. Justices can use a variety of signal words to articulate a particular posture toward a precedent, but the *Lemon* examples demonstrate the variety of ways in which the interpretive posture toward a precedent is a misleading characterization of a posture toward a doctrine. The Court can distinguish a precedent on the basis of a fact pattern and still utilize the legal doctrine from that precedent. The Court can cite a legal doctrine and utilize it in an opinion while not including strong enough language to trigger a follow coding. The Court can utilize a legal doctrine while including language that appears to weaken the doctrine, but the doctrine continues to be applicable nonetheless.

The key theoretical takeaway is that limiting or expanding the scope of a particular precedent is not the same conceptual matter as limiting or expanding the scope of a legal doctrine. A precedent is a particular application of a legal doctrine but is not the doctrine in and of itself. Studying the two concepts as one leads to mischaracterizing instances in which the Court utilizes a legal doctrine while taking a neutral or negative posture toward the precedents that produced that doctrine.

Methodologically, this means that tools that have been used to study the interpretation of precedent are inadequate for understanding the development and utilization of legal doctrine. While the reliability and validity of *Shepard's* has been supported in previous works (Hansford

and Spriggs 2006; Spriggs and Hansford 2000), I argue that *Shepard's* is only reliable and valid for the purpose for which it was created; that is, the interpretation of a particular combination of case facts and decision rules. However, what ultimately impacts the development of the law is the development of decision rules, and thus a method is needed that isolates the doctrinal content of court opinions. In the next chapter, I develop and evaluate a method for tracking the development of decision rules using automated content analysis.

CHAPTER THREE: MEASURING DOCTRINAL LANGUAGE USING AUTOMATED CONTENT ANALYSIS

In the previous chapter, I concluded that the development of legal doctrine and the interpretation of precedent are two conceptually distinct matters and thus should be studied in empirically distinct ways. A precedent is a particular application of a legal doctrine but is not the doctrine in and of itself. I demonstrated the theoretical and methodological consequences of collapsing these two concepts. Through a detailed case study of Establishment Clause jurisprudence, I also showed that *Shepard's Citations*, while a useful tool for understanding the interpretation of precedent, presents significant issues when used to analyze the development of doctrine. *Shepard's* depends on courts taking interpretive postures toward a particular combination of case facts, legal doctrines, and case outcomes. However, as these components are conceptually distinct, courts can and do utilize legal doctrine in situations that do not align with the precedents that established that doctrine.

Understanding the development of legal doctrine at this level of nuance presents a methodological issue. The use of a case study as a way of critiquing *Shepard's* creates the very generalizability concerns that Hansford and Spriggs sought to avoid by emphasizing the study of precedent vitality. Thus, the methodological challenge of this chapter is developing a generalizable measurement of the application of legal doctrine that is not dependent upon a case study methodology. To accomplish this, I propose a new measure known as doctrinal vitality and evaluate various automated content analysis tools as instruments for measuring doctrinal vitality.

Doctrinal Vitality vs. Precedent Vitality

Isolating the doctrinal content of opinions is an attempt to improve upon the methodology of Hansford and Spriggs by addressing two potential criticisms of their work that the authors anticipated. The first criticism is that the attempt to utilize a generalizable measure of the interpretation of precedent glosses over the nuanced ways in which the law evolves on a case-by-case basis. They noted that they "trade[d] off the ability to capture nuances in the content of the law as it changes over time for the ability to rigorously test the predictions from our theoretical model" (Hansford and Spriggs 2006, 15). This is certainly a fair point, and it is worth noting that my critique of *Shepard's* found in the previous chapter is dependent upon nuanced case studies that do not permit generalizability (exactly as they suggested).

What is needed is a generalizable way to measure the doctrinal content of opinions, which is the methodological goal of this chapter. This goal addresses a second potential criticism that the authors noted; that is, the possibility that "some decisions may create multiple legal rules" (Hansford and Spriggs 2006, 43n1). In instances where the Court positively interprets a precedent that contains multiple legal rules, it is not clear which rule (or combination of rules) the Court is positively interpreting. They controlled for the impact of multiple rules in their model by including a "breadth of precedent" variable that captured the number of specific legal provisions or issues that a case addresses. The presumption of this control is that cases that address multiple issues are more likely to utilize multiple decision rules. While this control (imperfectly) captures the extent to which the presence of multiple legal rules may affect the interpretation of precedent, it still fails to capture the interpretation of particular legal rules as a component of that precedent. To accomplish that goal, I introduce the concept and measure of doctrinal vitality.

Doctrinal vitality (to modify the concept of Hansford and Spriggs) captures the extent to which a legal doctrine has been utilized in previous opinions, independent of whether those decisions align with a set of case facts (and thus, independent of precedent citation). I model doctrinal vitality as a running average of the utilization of a doctrine in prior court opinions. Much like with precedent vitality, I theorize that the past utilization of a doctrine affects the future utilization of the doctrine. Broadly speaking, the utilization of a doctrine is a measure of the extent to which an opinion articulates and discusses that doctrine. An increase in the discussion of a doctrine over time increases its vitality, as the doctrine becomes more relevant to the development of the law. Conversely, a decrease in doctrinal discussion decreases doctrinal vitality, as the doctrine loses relevance. Given the theory of doctrinal development articulated in the previous chapter, all doctrines are expected to naturally depreciate as they age, although this relationship is attenuated by the ideological alignment of the Court and the vitality of the doctrine. Since the importance of precedents ultimately lies in their ability to produce legal rules and standards for future cases, the ability to extract and measure the doctrinal content of an opinion improves our understanding of the development and application of the law.

I am not arguing that the utilization of a doctrine indicates that the doctrine is the controlling factor that determines a case outcome. Rather, I am simply arguing that the more discussion that an opinion devotes to a doctrine, the more likely it is that the doctrine is influential in the logic of the opinion. Consequently, doctrines that are discussed more frequently in precedential opinions are more likely to remain influential in future opinions. A justice who desires to weaken a doctrine for ideological reasons can simply choose to not devote much discussion to the doctrine in the opinion. This is the logic behind the "distinguishing" behavior coded by *Shepard's* and critiqued in the previous chapter. Since the natural lifespan of a doctrine

is one of depreciation, justices do not have to actively weaken a doctrine through negative discussion of it; they can simply "distinguish" the doctrine into irrelevance and replace it with a new legal rule. In order to avoid doctrinal depreciation, the vitality of a doctrine must be actively sustained, just like precedents must be actively cited and positively interpreted to increase the probability of citation to the precedent in the future.

To show how the Court can simply ignore a doctrine while not actively working to weaken it, consider two examples from Establishment Clause cases discussed in the previous chapter. One of the examples, *Van Orden* v. *Perry,* was coded as "Distinguished" by *Shepard's*. In *Van Orden*, the Court upheld a Ten Commandments monument located on the grounds of the state capitol of Texas and did so by claiming that the *Lemon* test was irrelevant for the "sort of passive monument" at issue in the case (2005, 686). Significantly, the Court took no position on the overall validity of the *Lemon* test and offered no analysis of strengths and weaknesses of the test. The justices simply argued the test was not relevant. What made the *Van Orden* decision so intriguing is that the Court issued an opinion on the same day that invalidated the display of the Ten Commandments in courthouses and schools and used the secular purpose test as the decision rule.[13] The decision to utilize the *Lemon* test in one case and abandon it in the other appeared ideologically driven. *Van Orden* was a 5-4 decision with four of the five most conservative justices on the Court in the majority; *McCreary County* was a 5-4 decision with four of the five most liberal justices on the Court in the majority. To produce their preferred policy outcome, the justices found a way to modify which decision rule was applied.

[13] However, as observed in the previous chapter, this companion case (*McCreary County* v. *ACLU*) is not coded as "following" *Lemon* by *Shepard's Citations*.

A second example of ignoring the *Lemon* test is an extremely curious one, as the decision to ignore the test occurred in an opinion authored by the author of the test, Chief Justice Burger![14] In *Marsh* v. *Chambers*, the majority opinion acknowledged that the circuit court used the *Lemon* test to rule that legislative prayer led by state-employed chaplains was unconstitutional. However, Chief Justice Burger then proceeded to completely ignore the existence of the *Lemon* test, as the Court found legislative prayer constitutional on the basis of the history of the practice.[15] In Justice Brennan's dissent, he sharply critiqued the majority's refusal to apply the test, arguing that he had "no doubt that, if any group of law students were asked to apply the principles of *Lemon* to the question of legislative prayer, they would nearly unanimously find the practice to be unconstitutional" (*Marsh* v. *Chambers* 1983, 801).

While Burger's refusal to apply the *Lemon* test seems exceptional on its face (as he authored the test), the *Marsh* example points to the need to examine doctrinal language rather than precedent citation patterns to best understand the development of the law. Although the *Marsh* opinion did not contain the explicit language needed to trigger a "distinguished" coding by *Shepard's*, it served the same functional purpose, in that it limited the area in which the *Lemon* test applied. The lack of discussion of the test weakened the vitality of the test. This phenomenon is only apparent when studying the language of opinions. Recall from the previous chapter that one of the school prayer cases, *Lee* v. *Weisman*, called into question the future

[14] It should be noted that Burger later came to regret the trajectory of the *Lemon* test, as he felt that the Court was too rigidly applying the test rather than treating it as general guidelines. Note his dissent in *Wallace* v. *Jaffree*: "The Court's extended treatment of the 'test' of *Lemon* v. *Kurtzman*, suggests a naïve pre-occupation with an easy, bright-line approach for addressing constitutional issues. We have repeatedly cautioned that *Lemon* did not establish a rigid caliper capable of resolving every Establishment Clause issue, but that it sought only to provide 'signposts…' [O]ur responsibility is not to apply tidy formulas by rote; our duty is to determine whether the statute or practice at issue is a step toward establishing a state religion." (1985, 56).
[15] *Marsh* was decided in 1983, two years prior to Burger's dissent in *Wallace*.

validity of the *Lemon* test precisely because the Court declined to overrule the test while failing to apply the prongs in the course of the opinion (and possibly creating a new coercion test in the process). Measuring doctrinal language captures some of these nuances in a more generalizable fashion that is not dependent on the close reading of case study analysis.

Automated Content Analysis and the Measurement of Legal Doctrine

Doctrine is ultimately expressed in the language of opinions. Scholars of judicial politics have begun using the tools of computer-assisted content analysis (CACA) to study the language of opinions. Machine learning has been used to accurately classify the ideological direction of amicus curiae briefs (Evans et al. 2007). Evans and his co-authors also found that CACA could accurately identify the longitudinal reliance on certain ideological word groupings. Machine learning can be used to analyze other opinion features, such as opinion complexity (Owens and Wedeking 2011). Corley (2008) determined the extent to which the Supreme Court utilizes the language of party briefs in opinion writing (see also Oldfather, Bockhorst, and Dimmer 2012). Corley, Collins, and Calvin (2011) analyzed the incorporation of the language of lower court opinions into Supreme Court opinions. Collins, Corley, and Hamner (2015) identified similar linguistic patterns for amicus curiae briefs, although these briefs are used less frequently than party briefs or lower court opinions. Collins, Corley, and Hamner (2014) compared amicus briefs to other sources upon which opinions depend and found that amicus briefs do not tend to repeat information from other sources. These studies utilized plagiarism detection software to identify similarities, at least in terms of direct quotes.

To better capture the possibility of implicit or paraphrased influence, rather than direct quoting, some scholars have begun utilizing "bag of words" methods of unsupervised text analysis. These methods model the contents of documents as a collection of words, under the

assumption that similar documents will contain similar collections of words. Cosine similarity is one such method, which models the similarity of documents as a function of the frequency of terms in each document relative to the rarity of those terms in the entire corpus (known as a term frequency, inverse document frequency score, or TF-IDF). If two documents frequently use the same terms that are less frequently used in the larger corpus, then these documents are likely similar in content. Hinkle (2015) used this method to identify similar search-and-seizure cases on the circuit court level. Cosine similarity captures the similarity in the overall language of opinions; however, capturing the similarity in legal doctrine specifically requires isolating the doctrinal section from the case facts (e.g., Hinkle et al. 2012), which can be difficult in Supreme Court opinions.

Perhaps the most promising unsupervised method for quantifying doctrine is Latent Dirichlet Allocation. LDA assumes that each "item in a collection is modeled as a finite mixture over an underlying set of topics" and that "each topic…is modeled as an infinite mixture over an underlying set of topic probabilities" (Blei, Ng, and Jordan 2003, 993). Topic probabilities are identified by determining which words frequently appear together in a single document, as each word has a probability of being associated with a topic. To run an LDA model, the user specifies the number of topics expected to appear in a corpus of documents (through a trial-and-error process, as seen in model validation discussion below), and the model calculates the distribution of those topics in each document in the corpus and determines the words associated with each topic. Lauderdale and Clark (2014) used LDA to model the distribution of topics in opinions and mapped those topics onto issue areas from the Supreme Court Database. Since legal doctrines are expressed using a collection of words, I argue that LDA can be used to model a doctrine as one of the topics present in an opinion.

Jurisprudential Regimes as Test Cases for Doctrinal Modeling

The first step in testing the ability of content analysis tools to model the utilization of legal doctrine is to identify candidate doctrines to examine. Good candidate doctrines are expected to be influential in multiple cases across time (so that changes in doctrinal utilization can be examined) and have fairly programmatic articulations (so that doctrinal language can be easily identified). Candidate doctrines should also be tied to foundational precedents so that the origin point of the doctrine can be easily identified. Given these requirements, the jurisprudential regimes literature provides an excellent set of candidate doctrines. First proposed by Richards and Kritzer (2002), jurisprudential regimes are founded on controlling cases that establish two important facets of a legal doctrine: case factors and/or related levels of scrutiny or balancing. The theory posits a relatively long-lasting change in how the Court and individual justices apply these factors after the founding case(s). JRT does not discount the importance of attitudinal variables; instead, it represents a reconceptualization of the concept of precedent that attempts to rescue the legal model from a "strict and controversial" understanding of the influence of precedent (Gillman 2001, 477).

The original stream of JRT literature identified jurisprudential regimes in four policy areas: free expression law (Richards and Kritzer 2002), Establishment Clause jurisprudence (Kritzer and Richards 2003), search-and-seizure cases (Kritzer and Richards 2005), and administrative deference (Richards, Smith, and Kritzer 2006). To simplify the methodology: legal literature is used to identity a candidate regime cases or cases (e.g., the 1972 companion cases *Grayned* and *Mosley*) and resultant case factors and/or tests (e.g., content neutrality). If the factors/tests are statistically significant predictors of a change in voting behavior post-regime (relative to pre-regime), then a jurisprudential regime has been identified. While a significant

methodological debate developed around the ability to measure the impact that regimes have upon voting behavior (Bartels and O'Geen 2015; Kritzer and Richards 2010; Lax and Rader 2010a, 2010b; Pang et al. 2012), my goal is to utilize the doctrinal tests identified in the regimes literature as a tool for testing the ability to model the doctrinal components of opinions. For methodological validation purposes, I test two regimes: content neutrality and Establishment Clause jurisprudence.

The G*rayned*/*Mosley* free expression jurisprudential regime established a two-track pattern of analysis (Richards and Kritzer 2002). To be precise, *Mosley* articulated the content neutrality doctrine, whereas *Grayned* was a companion case that evaluated an almost identical fact pattern but also articulated a void-for-vagueness doctrine. Under the *Grayned/Mosley* regime, regulations that are found to be content-based are subject to strict scrutiny, which means that the regulation must be narrowly tailored to serve a compelling government interest. A regulation fails this test if the goal of the restriction is less than compelling or if another less restrictive regulation can be found to serve the same interest. Content-neutral regulations are subjected to a lower level of scrutiny known as intermediate scrutiny. These regulations, which fall in the category of time/place/manner restrictions, must be narrowly tailored to serve a significant government interest. In other words, content-neutral regulations are given more latitude to serve a broader range of government interests, due to the government showing no bias toward the content of the message. In a nutshell, the content-based track involves a balancing test between the restrictions placed on a speaker and the level of government interests being served (Stone 1983), whereas the content-neutral track prohibits viewpoint discrimination.

As an extended test of the content neutrality jurisprudential regime, Bartels and O'Geen (2015) evaluated the outcome of all Supreme Court cases that implicated free expression from

1946-2004. I started with the cases in their dataset but updated the case list through the 2016 term and expanded the analysis to all opinions issued for every case in the dataset.[16] The dataset contains 1,827 opinions across all opinion types. In this methodological validation section, my goal is to determine which of the three reviewed linguistic analysis methods provides the most validity for measuring the discussion of a legal doctrine. Measuring doctrine involves identifying key terms used to express a doctrine, and then evaluating the usage of those terms in opinions. I order my analyses in terms of expected validity of doctrinal identification; thus, I begin with cosine similarity, followed by plagiarism detection and then LDA.

Content Neutrality – Cosine Similarity

As noted, running a cosine similarity model involves reducing each document to a "bag of words," and then generating a TF-IDF score for each term in the document. Two documents are then compared based on the similarity of their TF-IDF scores. If a document is compared to itself, the comparison receives a score of 1; a comparison of two documents with no terms in common receives a score of 0. *Grayned* and *Mosley* have been identified as the companion precedents that established the content neutrality doctrinal regime.[17] Thus, if an opinion has a high cosine similarity score when compared with either *Grayned* or *Mosley*, it could theoretically share a high amount of doctrinal content. Without separating the case facts and legal doctrine sections of each opinion, the validity of cosine similarity as a proxy for shared doctrine is hampered (as the score will be biased in either direction by non-doctrinal terms).[18] One way to

[16] I included any cases from the Supreme Court Database that are coded as addressing the "First Amendment (Speech, Press, Assembly)" or "Protest Demonstrations," as these were the issues implicated by *Grayned* and *Mosley*. For the Supreme Court Database, see Spaeth et al. (2017).
[17] But see *supra* p. 46 about *Mosley* as the source of the doctrine; according to *Shepard's*, *Grayned* "follows" *Mosley*.
[18] As noted earlier, Hinkle et al. (2012) solved this problem by splitting opinions into law and fact sections. This splitting was done by manually reading opinions and tagging them with XML

investigate this potential validity issue is to manually evaluate a series of opinions that are the most like *Grayned* and *Mosley* and determine if doctrinal content is shared.

To compute cosine similarity scores for the opinions in my dataset, I used a Python script to process the text of the opinions and compare them with one another.[19] To reduce each document to a "bag of words," I removed all punctuation, numbers, stop words (commonly used words such as articles and conjunctions), and a user-defined set of common judicial terms that appear in the vast majority of opinions.[20] I also used the Natural Language Toolkit (Bird, Loper, and Klein 2009) to identify proper nouns and remove them and utilized the part-of-speech tagging function in conjunction with the lemmatization function to reduce all words to their dictionary root.[21] This ensured that there was not duplication of words that are only different because of their inflection or conjugation (but refer to the same idea). I then used Python to compute TF-IDF scores for every remaining term in each document and create a cosine similarity matrix for all documents in the dataset.[22] To ensure that the calculations were not biased by terms that appear in almost every opinion or terms that appear in almost no opinions, I limited the TF-IDF calculation to words that appear in 2%-98% of the opinions in the dataset. From the similarity matrix, I extracted the scores for each opinion and compared the opinions to

---

tags to allow the text analysis programs to identify the appropriate sections. Since I was aiming to automate the process as much as possible, I did not manually tag the opinions in this manner. Their analysis was also conducted on lower court opinions, which tend to have a more cleanly delineated law/fact distinction than Supreme Court opinions.

[19] All Python scripts are written in Python 3.6 and utilize the latest version of the relevant Python modules. My Python scripts are built upon a codebase authored by Joseph L. Smith, and I am indebted to his assistance.

[20] The list of words that were removed from each opinion can be found in the appendix.

[21] E.g., the verbs "run," "runs," and "ran" all reduce to their base form "run."

[22] Python scripts for all of the analyses in this study are available from the author upon request.

the majority opinion in *Mosley*.[23] Although *Mosley* is the point of comparison due to its utilization of the doctrine of interest, I am not looking for the extent to which an opinion copies from *Mosley* but the extent to which an opinion shares doctrinal language with it. Thus, I evaluated the similarity with all opinions, not just opinions that were issued after *Mosley*.

To check the validity of using cosine similarity to identify doctrinal similarity, I manually evaluated the twenty-five opinions that are the most linguistically similar to *Mosley*. Manually checking the most linguistically similar opinions is an appropriate validation method, as the likelihood of doctrinal utilization is expected to decrease as cosine similarity scores decrease (because of the decrease in shared language). To determine the presence of the content neutrality doctrine, I performed a manual search for key terms related to the doctrine.[24] If the terms were found, I analyzed the context of the terms to determine if the doctrine was utilized in the opinion. Doctrinal utilization means that the Court applies an aspect of the content neutrality doctrine to the logic of the case. Four of the top twenty-five opinions that are most like *Mosley* utilize the content neutrality doctrine, for a validity rate of 16%.[25] Sixteen of the top twenty-five opinions (64%) most like *Mosley* linguistically were actually decided prior to *Mosley,* and none of those opinions utilize the content neutrality language as *Mosley* does. Without a more fine-grained analysis, cosine similarity does not hold much promise for tracking doctrinal language in this legal area.

---

[23] I compared to *Mosley* only since *Mosley* established the language of the content neutrality doctrine. Doctrinal evaluation of *Grayned* would involve searching for the vagueness and overbreadth doctrines, which are not the doctrines of interest.

[24] I searched for "content," "content-based," "content-neutral," "neutrality," "interest," and "scrutiny." While this term list is certainly not exhaustive, some of these core terms should be present in some form in an opinion that utilizes the content neutrality doctrine.

[25] It should also be noted that the opinions in the dataset do not generally have high cosine similarity scores when compared to *Mosley*; only four cross the .5 threshold to establish similarity used by Hinkle (2015).

Content Neutrality – Plagiarism Detection

Plagiarism detection should allow for increased validity as a measure of doctrinal utilization by providing the ability to better supervise the meaning of the results (as the results are phrases, rather than a bag of words). To validate the plagiarism detection measure of doctrinal utilization, I employed the WCopyFind 4.1.5 software to compare all opinions in the dataset to *Mosley*. I set the software at the standard settings utilized in previous political science literature (match phrases of six words allowing for two imperfections; ignore nonwords, letter case, numbers, and outer punctuation). These settings were designed to prevent the matching of case citations. After comparing the documents, WCopyFind produces a count of the number of words that are perfect matches and a count of the number of words that are imperfect matches.[26] I again evaluated the twenty-five opinions that are the most similar to *Mosley* in terms of the highest number of words that are perfect matches with *Mosley*. The expectation is like cosine similarity: the more frequently an opinion shares language with *Mosley*, the more likely it is to utilize similar doctrine. To validate doctrinal utilization, I examined the reports produced by WCopyFind to determine if the common phrases were expressions of the content neutrality doctrine. Unlike with cosine similarity, I did not evaluate the entire opinion for utilization of the doctrine. The doctrine must have been expressed within the common phrases marked by the software; otherwise, the utilization of the doctrine was purely incidental to the results of the plagiarism analysis. Nineteen of the top twenty-five matches with *Mosley* utilize content neutrality or narrow tailoring doctrine (for a validity rate of 76%).[27]

---

[26] A perfect match is a phrase of at least six words that is the same as in the comparison opinion; an imperfect match contains at least seven words, as it is six perfect words with no more than two non-matching words.

[27] Recall that the narrow tailoring doctrine in the content context requires content-based restrictions to further a significant government interest.

Plagiarism detection clearly provides increased validity over cosine similarity for detecting doctrinal usage but does possess some potential weaknesses. The content neutrality regime is an ideal candidate for using plagiarism detection because expressing the doctrine often requires phrases long enough to be captured by the standard settings in the political science literature. Six words is the standard setting because five words or less may capture too many matches that are not truly borrowed language. Thus, if a doctrine can be expressed in fewer than six words (such as with the prongs of the *Lemon* test), plagiarism detection at the standard settings will fail to capture the doctrine. A second difficulty with plagiarism detection is the amount of manual validation required to ensure that the shared language is doctrinal in nature. A more resource-efficient method for detecting doctrine allows for more supervision of the language being analyzed. LDA is one such potential method.

Content Neutrality – Latent Dirichlet Allocation

As noted, this is a bag of words model that allows for increased supervision when compared to other models (because the words that comprise each topic can be evaluated). I preprocessed the opinions using the same steps described for the cosine similarity methodology, and then evaluated the opinions using LDA. Once the documents are preprocessed, the *textmining* Python package (Peccei 2010) was used to create a term-document matrix, which is a dictionary of all remaining words used in the corpus and their frequency in each document. The *lda* Python package (lda developers 2014) was then used to process this matrix to determine the topic composition of the corpus. This package uses collapsed Gibbs sampling (Griffiths and Steyvers 2004) with symmetric Dirichlet priors.

There are two Dirichlet priors: *alpha* refers to the document-topic density, and *eta* refers to the topic-word density. When the priors are symmetric, changing *alpha* results in changing the

number of topics that make up each document, and changing *eta* changes the number of words that make up each topic.[28] For this project, the symmetric priors are left at their defaults (*alpha* = .1, *eta* = .01). For replication purposes, *random_state* is set at one, and *iterations* is set at five hundred. As noted earlier, LDA requires a specification of the number of topics expected in the corpus. Various suggestions have been made for determining the appropriate number of topics by evaluating the semantic coherence of topics (Chang et al. 2009; Grimmer 2010; Mimno et al. 2011; Wallach et al. 2009). I followed the strategy of Grimmer in validating the number of topics – too few topics leads to unrelated words being clustered together, whereas too many topics leads to topic repetition (Grimmer 2010, 12). I evaluated topic coherence for twenty-five, thirty, and thirty-five topics across the corpus and determined that thirty topics provides the best balance of coherent topics.[29] I then identified which one of the thirty topics contained a collection of terms most closely associated with content neutrality and extracted the document-topic probability for that topic.[30] The document-topic probability measures the likelihood that an opinion contains the collection of terms most closely associated with the given topic.

As with the other methods, I evaluated the twenty-five opinions that have the highest document-topic probability for the content neutrality topic (excluding *Grayned* and *Mosley*, although they are included in the corpus when topics are created). Twenty-one of the twenty-five

---

[28] Wallach, Mimno, and McCallum (2009) argued that an asymmetric *alpha* is to be preferred as it discovers more precise topics; however, the LDA implementation used does not allow for asymmetric priors. Future work will compare the effect of an asymmetric *alpha* using an implementation such as Gensim, although Lauderdale and Clark (2014) found symmetric *alphas* to be appropriate for evaluating Supreme Court topics.

[29] The topics are included in the appendix, along with a descriptive label that identifies the doctrinal topic.

[30] As seen in the appendix, the 30-topic model identified a cluster of words that express the content neutrality doctrine, and a cluster of words that express the void-for-vagueness doctrine. Again, I limit the analysis only to content neutrality, but the ability of the model to find both doctrinal clusters is promising.

opinions that have the highest content neutrality topic scores explicitly utilize the language of the content neutrality doctrine. One of the remaining four opinions makes reference to content neutrality but only in the context of the lower court decision. Another of the remaining opinions makes reference to time, place, and manner restrictions, which are often discussed in conjunction with content neutrality. Of the remaining two opinions, one references the overbreadth doctrine from *Grayned*, and the other is the only opinion that contains no reference to doctrine from either *Grayned* or *Mosley*. Thus, of the identification methods evaluated, LDA carries the most promise for tracking the utilization of doctrine, in line with my expectations.

As a bag-of-words probability model, LDA does suffer from some weaknesses for my purposes. First, the bag-of-words strategy depends on words co-locating in a document, but not necessarily in the same context. Thus, while the doctrine may be present in a document if the words co-locate in the same document, a modeling strategy that looks for the words to co-locate in the same sentence may provide a more precise picture of doctrinal use. Secondly, while each topic does receive a probability score, this probability is assigned relative to every other identified topic in the corpus, even topics that do not appear in the document at all.[31] Thus, LDA models are best used to classify documents into their most likely topics, rather than as a precise measure of the proportion of a document dedicated to a topic.[32] Given these potential weaknesses, I turn to a second examination of the various text analysis methods using

[31] However, for topics that do not appear in the document at all, the probability scores are so close to 0 that the effect on other topic scores is minimal.

[32] Since LDA is traditionally used as a document classification model, many model validation suggestions involve "holding out" documents from the corpus, using the corpus to build the LDA model, and then testing the ability of the model to classify the held-out documents. I am not using LDA in this traditional classification manner, and thus do not perform this validation test; however, in evaluating the effectiveness of text analysis methods in Establishment Clause cases, I perform a check of the accuracy of the model's identification of the most likely topic.

Establishment Clause cases. As noted, the *Lemon* test prongs can be expressed in fewer words than content neutrality doctrine and thus provide a harder test for automated text analysis.

Establishment Clause – Text Analysis Validation

I begin with a validation of the cosine similarity comparison between all Establishment Clause opinions and *Lemon v. Kurtzman* (the original *Lemon* opinion).[33] Cosine similarity is much more successful in Establishment Clause cases. Twenty-four of the twenty-five opinions most like *Lemon* utilize at least one of the three prongs of the *Lemon* test, for a validity rate of 96%.[34] Plagiarism detection is equally as valid for the Establishment Clause as it was for content neutrality (nineteen of twenty-five for a rate of 76%). However, this validity statistic is somewhat deceptive, as the opinions with the highest rate of borrowed language are most likely to directly quote the three prongs as one unit from *Lemon*. The standard plagiarism detection settings still miss instances where the prongs are not quoted as one unit and are expressed in less than six words.

To add some theoretical backing to the number of topics selected for the LDA model, I set my baseline number of topics based on a manually-developed clustering of the factual areas covered by all majority opinions in the dataset.[35] Depending upon how broadly the opinions are clustered, I identified between ten and fifteen issue areas; thus, I evaluated LDA models using ten, fifteen, and twenty topics. Following Grimmer's validation suggestions, fifteen topics was

[33] I identified these cases by using the Supreme Court Database, specifically looking for cases where the issue was "First Amendment: establishment of religion" (for both parochial aid and non-parochial aid cases) or the legal provision was "First Amendment: establishment of religion."

[34] This validity is a function of a high amount of shared vocabulary. Unlike with content neutrality, all twenty-five opinions have cosine similarity scores that are .4862 or higher (twenty-one of twenty-five are .5 or higher).

[35] I list these broad clusters in the appendix. As an example, several cases deal with financing various aspects of parochial schools, so I consider that to be a topic cluster.

the preferred model.[36] While the fifteen-topic model did not identify all of my broad clusters, it identified several matches and a cluster of words associated with the *Lemon* test. As a post hoc validation of the fifteen-topic model, I compared the model's identification of the most likely topic for each majority opinion to my manual identification of the topic area for each majority opinion. The LDA model correctly identified the primary topical area of the majority opinion in 65% of cases. In cases where the primary topic identified was the *Lemon* test, I noted this as the appropriate identification if the opinion incorporated the *Lemon* test. In cases where the primary topic identified was a collection of words that were too general or incoherent to identify a specific topic area, I did not count this as a correct identification, even if the topic noted words that frequently appeared in the opinion. Thus, my validation method undercounted the accuracy of LDA classification for some opinions.

For these reasons, I used Python to conduct a quasi-supervised text analysis that looked for the presence of the prongs of the *Lemon* test on a sentence-by-sentence basis. I identified the stems of the key words most likely to appear together in a sentence that is referencing each prong of the test. The program then produced a count of the number of sentences that reference at least one prong of the test.[37] To validate this quasi-supervised count measure, I manually read half of the opinions in the dataset (126 out of 252 unique opinions) to see how accurately the program identified sentences that used *Lemon* test terms. Although no biases were expected, I chronologically sorted all of the opinions by LexisNexis citation and read the oldest 25% and newest 25%, across all opinion types. The average percentage of agreement between my count of

[36] I list the term clusters that LDA identified in the appendix. The appendix also contains replication settings for the LDA model.
[37] For the entanglement prong, the sentence must contain "entangl." For the purpose prong, the sentence must contain "secular" and "purpose." For the effect prong, the sentence must contain "effect(s)" and "religio." Words were stemmed and pluralized as written to capture variants.

Lemon test sentences and the program's count for each opinion was 82.92%. 95% of the disagreements were by were four sentences or less. One outlier pre-*Lemon* opinion had a very inaccurate count due to multiple uses of the word "entanglement" in a completely different context; however, many of the inaccuracies in other opinions appeared to be related to differences in where I manually split a sentence and where the program did.[38] This splitting disagreement does not appear to be a product of any particular justice's writing style or era and thus is not a measurement bias issue.

My validation rules were quite strict. As an example, the program counted this sentence from Brennan's concurrence in *Walz* as containing a reference to the purpose prong: "To the extent that the exemptions further secular ends, they do not advance 'essentially religious purposes'" (1970, 693). I did not count this sentence in my manual validation because "secular" is not modifying "purpose." However, the sentence can be read as referencing the purpose prong due to the synonym "ends" and the reference to "religious purposes." In future iterations of this project, I intend to expand the count and validation to account for situations like this; here, I wanted to test the method for accuracy using the most restrictive means possible.[39]

I then used the supervised count measure to validate the unsupervised topic model. The two measures have a weak positive correlation (r = .3583); as noted, the measures are capturing two different types of linguistic structures, so this weakness of correlation is unsurprising.[40]

---

[38] The Python script split an opinion into sentences using the Natural Language Toolkit, which predicts sentence splits on the basis of punctuation. Its ability to accurately split sentences can be affected by ellipses and legal citations. The sentence tokenizer can be customized as to how it deals with certain punctuation patterns, but for this study, I left it at its default settings.

[39] The program did not count any synonyms, although justices occasionally used words like "sectarian" for "religious" or "nonreligious" for "secular." Accounting for those broader synonyms is part of the next iteration of project.

[40] I calculated a proportional variable as an additional validation of the unsupervised topic model (since the LDA probability score is a proportion). The correlation between the *Lemon* topic

Difference of means testing confirmed that the two measures are broadly capturing usage of the *Lemon* test. If the sample is split in half at the median LDA probability score for the *Lemon* topic, the mean number of sentences that utilize the *Lemon* prongs is significantly higher (p = .0000) above the median when compared to below the median (this is also true if the *t*-test is limited only to my manual validation set).[41] While the two measures do not correlate well in terms of a linear relationship, both measures broadly capture increasing reference to the *Lemon* test in an opinion.[42] As an additional validation, I checked each opinion in the dataset to see if it used one of the prongs of the *Lemon* test to contribute to the argument of the opinion. This validation process is separate from confirming that the program accurately counts mentions of the *Lemon* test; in 93.22% of opinions where the program noted at least one sentence that includes *Lemon* test terms, I manually determined that the opinion utilized *Lemon* doctrine as part of its argument. I then ran logistic regressions with the use of the *Lemon* prongs as the dependent variable, and the raw count of *Lemon* sentences and the LDA probability scores as the sole independent variable in the respective models. The results of the validation models are shown in table 2.

probability and *Lemon* sentence count proportion is .4999, indicating a moderate correlation. Sentence counts were calculated using the Natural Language Toolkit.

[41] For the full set of opinions, the mean number of *Lemon* test sentences below the LDA score median was 2.87 and above the median was 9.18. For the manual validation set, the mean below the LDA median was 2.3 and above the median was 9.43.

[42] Using LDA for this purpose is best seen as a broad measure of utilization due to the fact that the topic that includes *Lemon* test terms also captures other unrelated words. The sentence count focuses more narrowly on *Lemon* test usage.

Table 2. LDA and Sentence Count Validation Models (*Lemon* Test)

Variable	***Lemon* Prongs Usage**	***Lemon* Prongs Usage**
LDA Probability Score	8.032 (1.788)***	-------
Raw Sentence Counts	---------	1.574 (.229)***
N	263	263
LR X^2	25.46***	238.71***

* $p <= .05$ ** $p <= .01$ *** $p <= .001$. Robust standard errors in parentheses.

In both validation models, the raw counts and topic probabilities, respectively, are highly significant predictors of *Lemon* doctrine usage. In other words, as the topic probability (or sentence counts) for the prongs of the *Lemon* test increase, the probability that the opinion utilizes the *Lemon* test as a legal doctrine also increases. Thus, both measures are good indicators of doctrinal utilization, but both measures have unique advantages and disadvantages.

The LDA topic score measure has the virtue of being unsupervised, in that I did not specify any particular combination of words for the program to find. This allows for the model to find a variety of combinations of language associated with the legal doctrine, which results in the ability to detect a broader range of linguistic influence. However, this dependence upon a "bag of words" modeling strategy does have the drawback of the topics capturing words unrelated to the doctrinal area within the same topic as doctrinal words; thus, the probability scores are a rough estimate of doctrinal utilization. While the actual construction of topics is an unsupervised process, model specifications (specifically, the number of topics) does require a fair amount of guesswork and post hoc evaluation, which raises some questions about facial validity.

The raw sentence count measure addresses some of these facial validity questions by allowing for greater control over the language that is identified as co-locating. By requiring the terms to co-locate in the same sentence, the measure paints a more precise picture of the way in which language operates, rather than analyzing the co-location of terms as a decontextualized

"bag of words." The measure also does not depend on the guesswork and post hoc analysis of topic composition that LDA does. However, there is an implementation trade-off, in that the model only identifies terms that I specify; thus, it has no freedom to find broader linguistic influences. Given the advantages and disadvantages of both measures, it is useful to proceed by using the measures in tandem to analyze doctrinal utilization.

Conclusion

This chapter has introduced the theoretical concept of doctrinal vitality and explored methodological tools for measuring the utilization of legal doctrine. I distinguished doctrinal vitality from precedent vitality and introduced broad theoretical expectations for how the vitality of a doctrine affects its future use by the courts. Using free expression and Establishment Clause jurisprudence as test cases, I demonstrated the usefulness of various text analysis tools for analyzing legal doctrine, specifically determining that the unsupervised topic modeling method known as Latent Dirichlet Allocation and the supervised method of automated sentence counting holds promise for tracking the utilization of legal doctrine. In the next chapter, I apply those tools to the substantive question of the utilization of doctrine by the Supreme Court.

CHAPTER FOUR: DOCTRINAL UTILIZATION IN SUPREME COURT OPINIONS

The primary focus of this chapter is the measurement of the doctrinal components of Supreme Court opinions and an analysis of the factors that influence the decision to utilize a particular legal doctrine. To frame theoretical expectations for the utilization of doctrine, I begin with a review of how precedent age, judicial ideology, and precedent vitality affect the decision to cite precedents. I use these findings about precedent citation to argue for my theoretical expectations regarding the utilization of legal doctrine in Supreme Court opinions. I then utilize Latent Dirichlet Allocation topic modeling and automated sentence counting to test my expectations in three legal areas: content neutrality, Establishment Clause, and employment discrimination. I conclude with a generalized analysis of my findings, including a brief discussion of avenues for future research.

Precedent Age and Citation Patterns

One intuitive implication of the theory of doctrinal development that I articulated in chapter two is that precedents possess a natural lifespan. As the courts engage in the process of defining the boundaries of the case space to which a precedent applies, older precedents begin to lose relevance (and are cited less frequently). Landes and Posner (1976) called this phenomenon the "half-life" of a precedent. In their groundbreaking analysis of citation behavior on the Supreme Court during the 1974-75 term, the authors found that the median age of a cited precedent was 9.8 years. Using economic theory, the authors argued that precedents provide a stock of "legal capital" that naturally "depreciates" over time. Some types of precedents

depreciate more rapidly than others. Precedents that cover a narrowly focused fact pattern depreciate more rapidly than general precedents, because narrowly focused fact patterns are more likely to be "rendered obsolete by a change in the social or legal environment in which the precedent is applied" (Landes and Posner 1976, 268). This finding echoes the doctrinal lifespan theory of Levi, in that the law must evolve to adapt to changes in the sociological or technical environments.

Landes and Posner admitted that one of the limitations of their study was the small sample size, and they encouraged future researchers to examine the relationship between age and citation over a larger sample. A wide swath of literature has continued to find support for the idea that older precedents are cited less frequently. Black and Spriggs summed up the findings nicely: "It is now a well-established regularity that Supreme Court precedents experience depreciation, whereby cases are generally less likely to be cited as they become older" (2013, 4). Black and Spriggs made an important contribution to this literature by empirically examining various factors that affect the depreciation rate of a precedent, including the generality of the precedent and the ideology of the justices. Precedent depreciation rates can also be impacted by the legal importance of a precedent. Fowler and Jeon (2008) found that the most legally important cases (as identified by human experts and by the location of the cases in citation networks) tend to "peak" in terms of their citation frequency around twenty-five years from when the case is decided, before entering into the depreciation period. Thus, more legally important cases undergo a much slower development and decline period.

One notable study that does not treat precedent age as a substantive variable that explains citation patterns is Hansford and Spriggs (2006). While the authors include precedent age as a control variable, they reject age as a substantive explanation because of difficulties with

theorizing the impact that age could have upon precedent citation. While it is intuitive to argue that precedents naturally depreciate over time, it is also possible that the passage of time strengthens (and thus "institutionalizes") certain precedents. Given this institutionalization possibility, Hansford and Spriggs argued that is not possible to offer a testable directional hypothesis about the impact of precedent age, at least not without interacting it with other variables such as precedent importance. The need for this interaction implies that age is not the factor that is shaping the depreciation of precedent; rather, it is the vitality of a precedent that impacts its citation frequency. This theory of precedent vitality also allowed Hansford and Spriggs to speak to variables that reveal the "politics of judging" rather than variables that cannot be "directly influenced by the behavior of judges" (2006, 24). However, when Hansford and Spriggs included precedent age as a control variable in their model, they found older precedents are less likely to be interpreted at all, as expected.

These findings about precedent age, particularly as it relates to the legal importance of precedents, fit well with my expectations for the lifespan of a legal doctrine. I contend that doctrinal age decreases the likelihood of utilizing the doctrine, primarily due to the way that doctrine evolves as it is communicated through a chain of precedents. For Levi (1948), doctrinal breakdown occurs when the boundaries of the case space lose their usefulness for distinguishing case facts. The boundaries of a zone become increasingly clear over time, making it easier to see that a new set of case facts do not fit into that zone; thus, a new zone (and consequently, new doctrine), must be created. However, since doctrinal creation is often a microevolutionary process (Cardozo 1921), this new doctrine will be legitimized using more recent adaptions of the existing doctrine, rather than returning to earlier formulations of the doctrine. Thinking in terms of clarifying the boundaries of a doctrine offers a different way to consider the theoretical

implications of generalized doctrines depreciating more slowly. The more general a doctrine, the longer it takes to dispel the ambiguity surrounding the application and interpretation of the doctrine. More narrowly applied doctrines will be clearer (and less broadly applicable) from the outset of their lifespan, and thus more likely to depreciate quickly.

This argument is similar to Landes and Posner's contention that older precedents are less salient to contemporary legal questions, but also draws from Lindquist and Cross's (2005) critique of Dworkin's chain novel hypothesis. Dworkin theorized that the communicative function of precedent is like a chain novel, in which each justice or judge is increasingly constrained by the stream of precedents that has gone before. Lindquist and Cross found that the reverse is the case; more precedents create more options for articulating one's preferred policy position, not less. In other words, the accumulation of precedents in a legal area creates a buffet of options rather than a constraining chain novel. In evolutionary terms, this widening variety of options causes a doctrine to slowly evolve into a new doctrine rather than retaining the language of the original doctrine. This expectation provides another explanation for the natural depreciation of precedents over time: more recent adaptions of a doctrine are found in newer precedents; thus, newer precedents are more likely to be cited.

Judicial Ideology and Precedent Citation Patterns

Lindquist and Cross's finding about the accumulation of precedents providing a range of options by which a justice can actualize policy preferences is reminiscent of the Segal/Spaeth critique of the legal model. Under their critique, stare decisis is not constraining because the justices possess a multitude of options by which they can appear to be abiding by precedent when they are simply using precedents as "window dressing" for their ideological preferences. The strategic model articulated by Epstein and Knight (1998) suggests that the relationship

between ideology and precedent citation is more complex than the "window dressing" metaphor. Under their model, justices have ideological preferences and are constrained by the normative expectations of stare decisis. Epstein and Knight noted that "justices who wish to establish policy that will govern the future activity of the society in which their Court exists will be constrained to choose from among the set of rules (precedent and the like) that members of that society will recognize and accept" (1998, 45). Analyzing bargaining memos, they found that justices exhibited a significant concern with how precedents are interpreted (in nearly one-quarter of memos examined). Justices work within the precedential framework available to them and attempt to shape the meaning of that framework going forward.

The strategic model of Epstein and Knight is part of a larger theoretical tradition that sees justices as "single-minded seekers of legal policy" (George and Epstein 1990, 325). Thinkers in this tradition attempt to unify the outcome-oriented concerns of the attitudinal model with the stare decisis expectations of the legal model but do so in a way that places a greater emphasis on the content of legal rules than traditional attitudinalists do. This distinction is captured by the "Outcome Postulate", which states that "justices prefer Court opinions and legal rules that reflect their policy preferences" (Maltzman, Spriggs, and Walhbeck 2000, 17). The key insight of this literature is that justices are concerned with more than just case outcomes, against Segal and Spaeth. Justices are also concerned with how the legal rules contained in precedents shape the future direction of the law, with the goal of making the future actualization of their policy preferences more likely. Given this goal, precedent citation becomes a more strategic process than the simple selection of a precedent that justifies a particular case outcome. Justices must consider the doctrinal content of precedents and the interpretive status of the precedents in the

case law of the Court to select the ideal precedents that allow for preferred future legal development.

This interaction between ideology and precedent strength is the key insight of Hansford and Spriggs in their articulation of precedent vitality. Broadly speaking, Hansford and Spriggs confirmed the intuitive expectation that a court that is closer to the majority coalition of a precedent, ideologically speaking, is more likely to positively interpret the precedent (and vice versa, see Hansford and Spriggs 2006, 64, 83). Black and Spriggs (2013) also found that ideological distance is positively correlated with precedent depreciation over time; in other words, an increase in ideological distance from the median of the precedential court is positively correlated with a decrease in precedent citation. However, Hansford and Spriggs found that this relationship is attenuated by the vitality of a precedent. They defined precedent vitality as the extent to which a precedent is more "legally authoritative than others" (Hansford and Spriggs 2006, 22). Mathematically, they operationalized this concept as a sum of the positive and negative interpretations that the Court applies to a precedent. Positive interpretations increase the vitality of a precedent; conversely, negative interpretations decrease the precedent's vitality.

As part of their goal of strategically creating a legal environment that promotes the actualization of their policy preferences, justices are particularly concerned with the vitality of legal precedents. The findings of Hansford and Spriggs support the notion that justices are not pure attitudinalists. While justices are obviously more apt to positively interpret precedents with which they ideologically resonate, the effect of ideological distance on positive interpretations is weakened for precedents that have strong vitality. More vital precedents are more constraining. However, justices who are more ideologically distant from a precedent would also like to weaken strong precedents if possible, given their policy goals. Thus, the higher the vitality of a

precedent, the greater effect that ideological distance has on the probability of negative interpretation. Since strong precedents are more likely to carry weight in the future, justices that disagree with those precedents feel more compelled to weaken them at any opportunity. Less vital precedents can be more easily ignored, as they are less likely to be influential anyways.

In chapter two, I argued that the doctrinal components of a precedent should be seen as a subset of the precedent. Drawing upon the "Outcome Postulate," I argue that it is the vitality of these legal rules and standards that actually concern justices, rather than the strength of a particular combination of legal doctrine and case facts. In terms of theoretical expectations regarding the relationships between age, ideology, and vitality, I draw from the work of Hansford and Spriggs. My innovation is methodological in nature, as I address the measurement issues that arise from dependence upon *Shepard's Citations* by isolating the doctrinal language in court opinions to measure the vitality of legal doctrines.

Hypotheses

The theory of doctrinal utilization that I have articulated suggests several testable hypotheses. As a reminder of my expectations, I list those hypotheses below, and then test each hypothesis across multiple areas of law.

$H_{4\text{-}1}$: As chronological distance from the precedent that established a doctrine increases, doctrinal utilization will decrease.

$H_{4\text{-}2}$: All else being equal, as doctrinal vitality decreases, doctrinal utilization in future opinions will decrease (and vice versa).

$H_{4\text{-}3}$: Doctrinal utilization will be higher in opinions in which the ideological preferences of the opinion's author or median coalition member align with the author or coalition of the original precedent that established the doctrine.

Content Neutrality – Doctrinal Utilization Models

In light of my methodological findings from chapter three, I modeled the utilization of the content neutrality doctrine using three different dependent variables: LDA topic probability score, raw sentence count, and a proportional measure that captures the percentage of sentences in each opinion that contain the language of the content neutrality doctrine. All three measures are capturing the same broad concept of doctrinal utilization in an opinion. The LDA topic probability score captures the probability that the topic associated with the doctrine appears in an opinion. The raw sentence count measure captures the number of sentences in the opinion that utilize a set of terms associated with the doctrine. The proportional measure captures the percentage of sentences in an opinion that utilize doctrinal language to account for differences in opinion length. For the LDA topic score, I used the specifications from the methodology tests in chapter three. For the raw sentence counts, a sentence must have contained the stems "content" and "neutral," "content" and "based," "government" and "interest," "tailor," "scrutin," or some combination of those stems to be counted as utilizing the content neutrality doctrine. The proportional measure used this count and the total number of sentences in the opinion as counted by the Natural Language Toolkit sentence tokenizer.

I began with two descriptive tests of $H_{4\text{-}1}$, the doctrinal age hypothesis. The first test was a *t*-test to evaluate the validity of using *Grayned* and *Mosley* as the reference opinions for the content neutrality regime. If the opinions are central in establishing the content neutrality doctrine, then the mean probability of an opinion utilizing the language of the doctrine should be significantly higher in opinions issued in 1972 or later than in opinions issued prior to 1972. This is supported by the evidence; the mean probability of utilizing content neutrality doctrinal language is .0440 post-1972, but only .0054 pre-1972 (*p*-value = .000 for a two-tailed test). This

relationship is also true for the sentence count measure (2.0668 post-1972, 0.1430 pre-1972, *p*-value .000) and the proportional measure (.0132 post-1972, .0014 pre-1972, *p*-value .000). The expected age relationship in $H_{4\text{-}1}$ can be visually tested using Lowess plots of the utilization measures for each opinion. I provide plots of each measure below, including one plot for all opinions and a second plot for majority opinions only (as majority opinions carry precedential weight that other opinion types do not). Each plot includes a reference line for the 1971 term, when *Grayned* and *Mosley* were decided (see figures 8, 9, and 10).

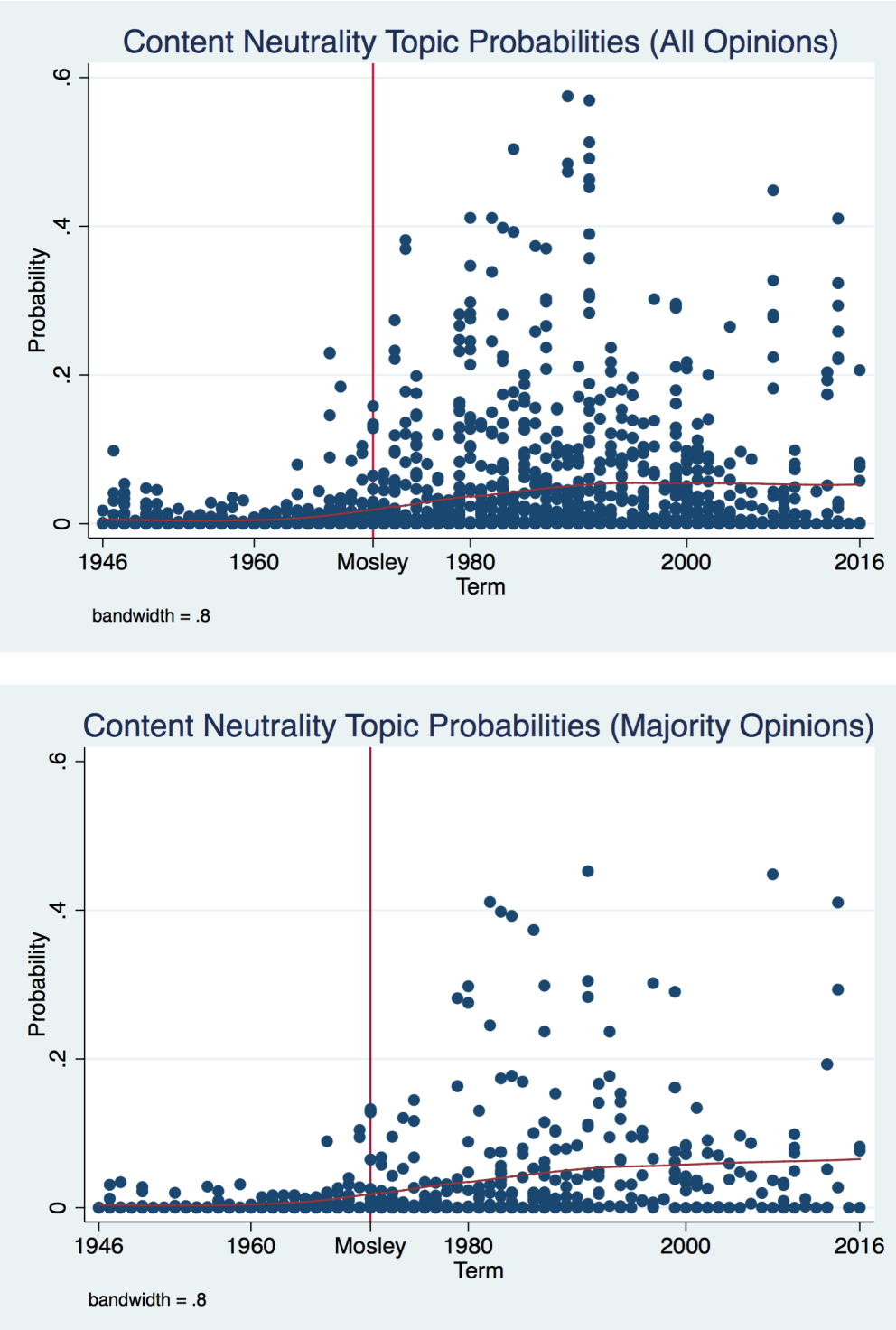
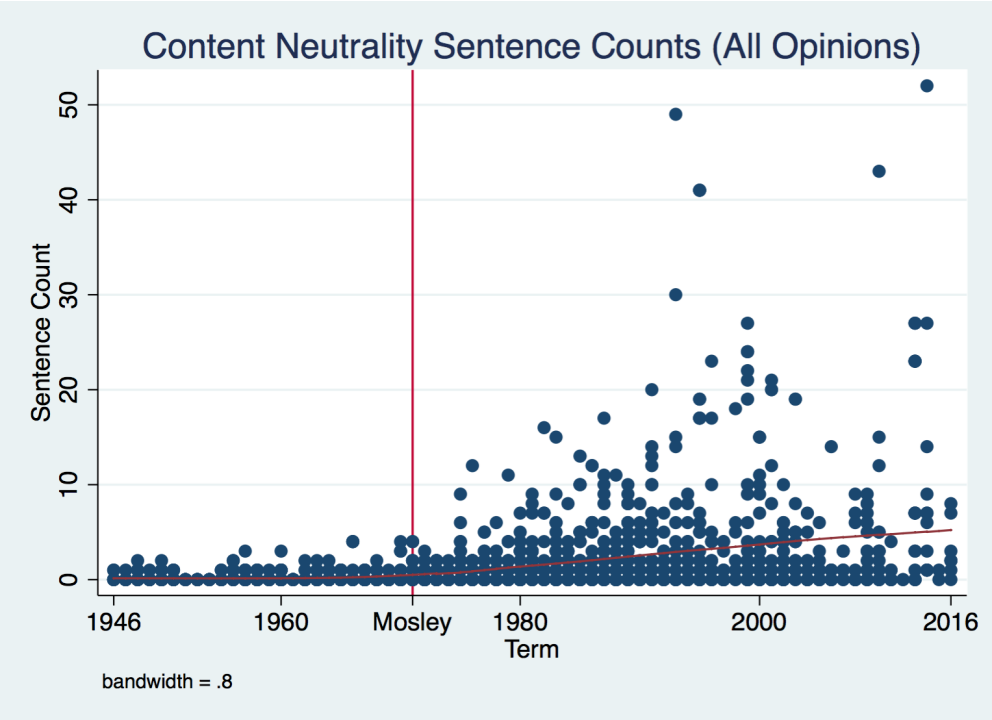


Figure 8. Content neutrality (LDA topic probability scores)



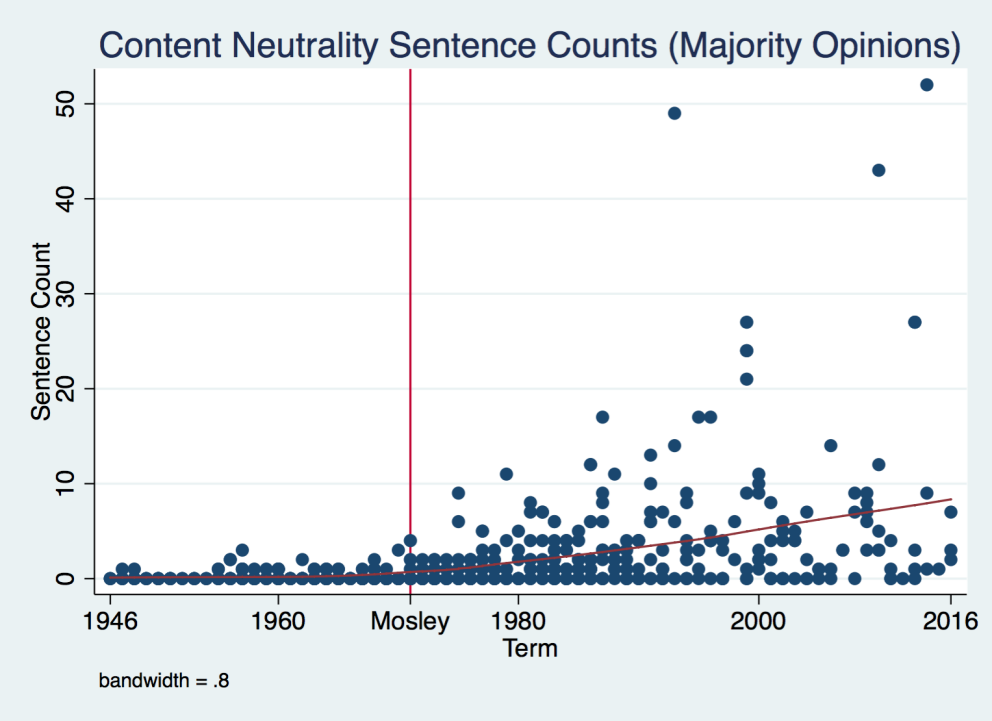
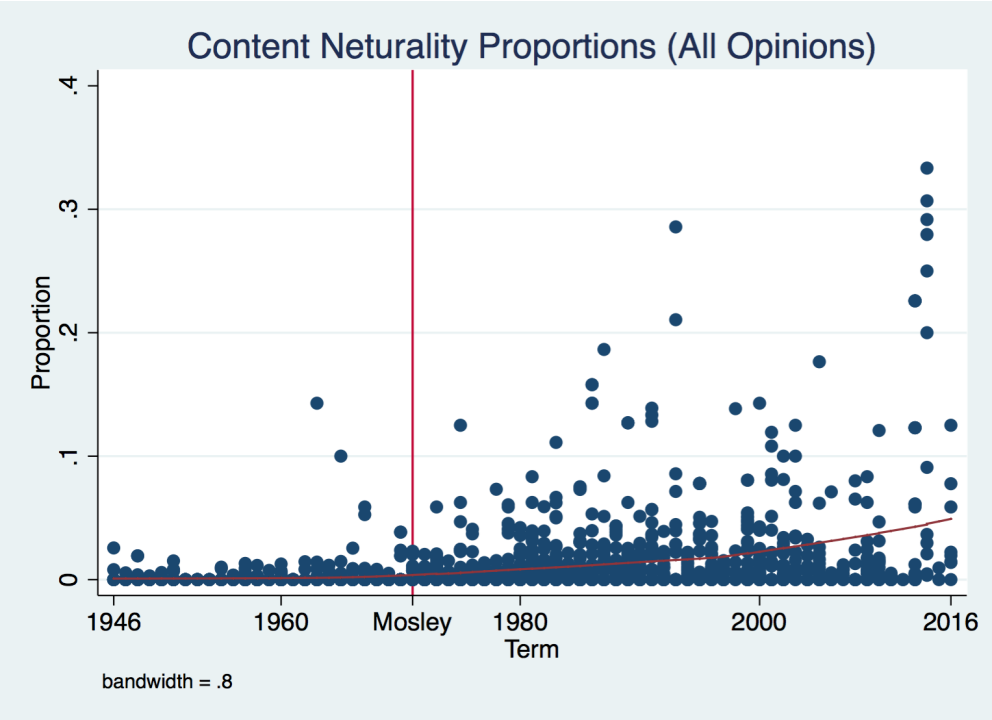


Figure 9. Content neutrality (sentence counts)



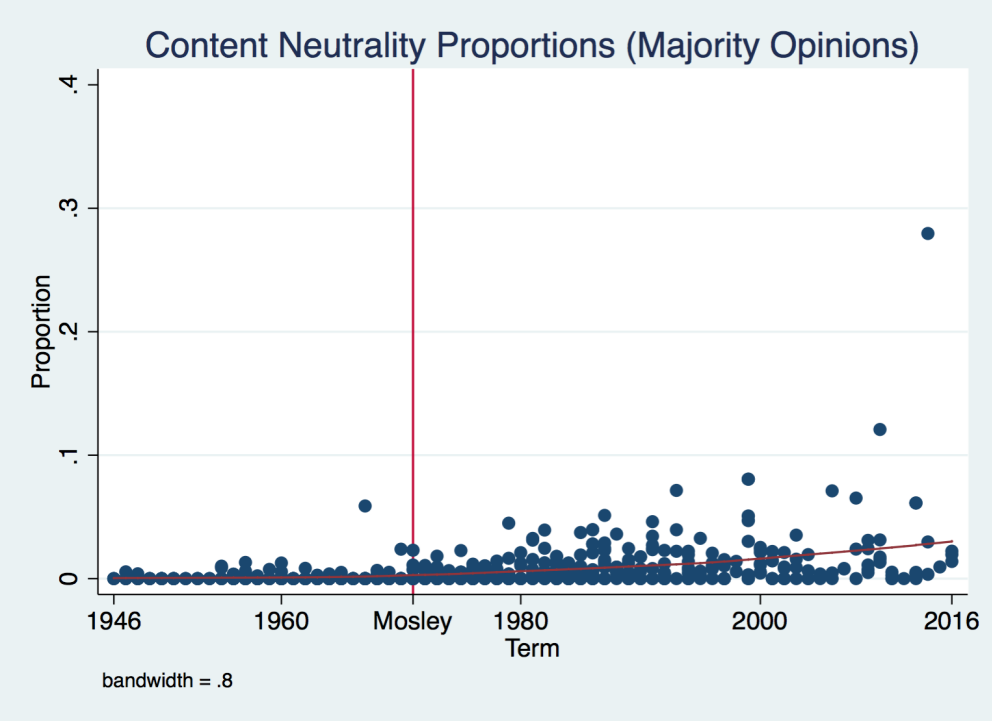


Figure 10. Content neutrality (sentence proportions)

These visual tests of $H_{4\text{-}1}$ appear to contradict my expectations for doctrinal lifespan.

Across all plots, the Lowess line rises over time, indicating an increase in doctrinal discussion

(although on the LDA plots, the slope of the line appears to gradually flatten). This could mean that the content neutrality doctrine does not exhibit the expected depreciation; it could also mean that the moment of depreciation has not yet been reached in the lifespan of the doctrine. The Lowess plots are not the most precise method for capturing the impact of age upon doctrinal depreciation. For that, I used regression modeling; however, it is encouraging to note the similarities between all plots, which indicates that my various ways of modeling doctrinal utilization are capturing similar concepts. To measure the age of the content neutrality doctrine, I calculated the chronological distance between an opinion and when *Mosley* was issued.[43]

To test $H_{4\text{-}2}$ (the vitality hypothesis), I modeled doctrinal vitality as a running average of doctrinal utilization in majority opinions decided in the previous five and previous ten case years.[44] A case year is a year in which a free expression case is decided. I calculated doctrinal vitality using only majority opinions since majority opinions are binding precedent upon future courts. This measure is inspired by the concept of precedent vitality from Hansford and Spriggs but does not attempt to capture the difference in positive and negative discussion of a doctrine. The assumption of $H_{4\text{-}2}$ is that if the doctrine has been frequently discussed in the immediately preceding case years, then justices will feel more obligated to discuss the doctrine in the current case year. I modeled doctrinal vitality to match the way that the dependent variable is measured

[43] Black and Spriggs (2013) used fractional polynomial modeling to determine the best quadratic specification for their doctrine age variable. Since the Lowess lines are not clearly linear, I ran preliminary tests using Stata's *fp* command to determine the best specification. In each results table, I report the best specification (measured by the smallest deviance). If the quadratic specifications do not produce models that are significantly different from the linear (null) model, I report doctrine age as a linear variable.

[44] At the beginning of the lifespan of a doctrine, there are obviously not five and ten years of case history from which to draw. In those instances, the average is calculated using whatever amount of case history is available.

(i.e. when the dependent variable is sentence counts, the vitality score in the model is an average of sentence counts).

To test $H_{4\text{-}3}$, I assigned an ideological score to the content neutrality doctrine by assuming that the doctrine reflects the ideology of the median member of the coalition that joined the majority opinion in *Mosley*. Hansford and Spriggs (2006) used this assumption when modeling ideological distance; it is supported by the findings of Carrubba et al. (2012). I made a similar assumption for the doctrine of each opinion in my dataset (for concurrences and dissents that are only joined by the author, I used the sole author's ideology). As ideological distance from the opinion author or opinion coalition increases, I expect doctrinal utilization to decrease. I measured justice ideology using two scores standard to the literature (Martin-Quinn and Segal-Cover), and a third score I created to measure ideology exclusively in content neutrality cases.[45] As noted by Lauderdale and Clark (2012), different issue areas have different median justices on the Supreme Court, implying that justice ideology is dependent upon the issue area. Since both Martin-Quinn scores and Segal-Cover scores are not issue-specific, I developed a free expression liberalism score that is the cumulative proportion of liberal votes for every case in the dataset (as coded by the Supreme Court Database).

For example, if a justice casts one liberal vote out of two free expression votes that were cast in the justice's first year in the dataset, then that justice's liberalism score is .5.[46] In the next term, if the justice casts three liberal votes out of four possible cases, then the liberalism score for that term increases to .6666 (as the justice has now cast liberal votes in four out of six

[45] For an explanation of Martin-Quinn scores, see Martin and Quinn (2002). For an explanation of Segal-Cover scores, see Segal and Cover (1989).
[46] Sometimes, a free expression case will present more than one legal issue. In these cases, this ideology score is the average of the votes across all issues.

possible cases). This allows the justice's free expression ideology to shift over time, ties that ideology to voting behavior, and is not circular because votes are being used to predict the utilization of doctrinal language rather than votes. After assigning all justices a liberalism score for each year, I calculated the absolute distance between the ideological score of each opinion's author and the score of Justice Marshall in the 1971 term, when he wrote *Mosley*. I made a similar calculation for the opinion coalition for each opinion and the opinion coalition in *Mosley*.[47] As an additional ideological variable, I included a dummy variable that specifies the direction of the vote associated with the opinion (either liberal or conservative, as per the SCDB). *Mosley* was a liberal outcome, so the expectation is that opinions written by justices voting in a liberal direction are more likely to utilize content neutrality doctrine.[48] I also included controls for justice and coalition ideology measured in absolute terms, not relative to Marshall and the *Mosley* coalition. As with the directional control variable, I expect that an increase in the liberalism score of the justice or coalition will lead to an increase in doctrinal utilization.

It is plausible that there are characteristics of justices outside of ideology that affect the decision to discuss a doctrine in an opinion. For instance, some justices may be more committed to the norm of stare decisis than other justices and thus may feel a greater obligation to ground their logic in the logic of previous opinions. Justices may possess some particular affinity (or distaste) for the content neutrality doctrine regardless of the justice's ideological predisposition in free expression cases. There could be writing style quirks that impact the ability to model the

[47] An opinion may encompass multiple cases due to the Court consolidating cases. Occasionally, justices will vote differently in different consolidated cases, leading to split opinion coalitions. I averaged ideology scores across all cases covered by a particular opinion to account for this possibility. If the opinion author voted differently in different consolidated cases, the opinion is included in the dataset multiple times to account for the different votes covered by the opinion.
[48] As I am using ideology to predict opinion language, rather than votes, there are no endogeneity issues, nor is there a need to lag the ideology scores to correct for endogeneity.

usage of language, such as the propensity to directly quote previous opinions versus paraphrasing those opinions with synonymous language. I modeled these non-ideological possibilities by including a fixed effects variable for each opinion author.[49]

I began by using the LDA topic score from the model validations (thirty topic model) in the previous chapter as my dependent variable for testing doctrinal utilization. Unlike in the language validation models, I only analyzed opinions that were issued concurrent to or after *Grayned/Mosley*, when the doctrine was established. A generalized linear model with a logistic link function from the binomial family and robust standard errors was appropriate since document-topic probabilities theoretically range from 0 to 1. The results are reported in table 3.

Table 3. GLM Results for Content Neutrality Topic Probabilities

Variable	***Grayned***	***Mosley***
Doctrine Age	.028 (.013)*	.027 (.013)*
Author Ideology	.575 (2.809)	.998 (2.796)
Coalition Ideology	-.688 (.464)	-.971 (.395)*
Author Ideological Distance	-.660 (2.960)	-.194 (2.954)
Coalition Ideological Distance	-.503 (.686)	-.819 (.630)
Liberal Opinion Direction	.066 (.123)	.064 (.126)
Justice Fixed Effects	Not significant	Not significant
Vitality (Five-Year Average)	4.059 (4.981)	3.889 (4.998)
Vitality (Ten-Year Average)	-5.328 (6.935)	-5.319 (6.919)
N	1267	1267
Wald X^2	97.11***	100.59***

* $p <= .05$ ** $p <= .01$ *** $p <= .001$. Robust standard errors in parentheses.

[49] Another way of thinking of the fixed effects variable is as an author-specific error term that captures variation not otherwise modeled through ideological effects.

I ran models for both *Grayned* and *Mosley* because of differences in the opinion coalitions that affect the measurement of ideology. The Supreme Court Database codes a case as having a special concurrence if the concurrence agrees with the result, but not the logic of the opinion. In my ideological calculations, I left justices out of the majority opinion coalition if they filed special concurrences, since their disagreement with the court's reasoning could involve disagreement with the legal rule. In *Mosley*, two conservative justices (Blackmun and Rehnquist) concurred in the result but did not provide a written special concurrence to note what they disagreed with in the logic of the opinion. Given that did they not file special concurrences in *Grayned* and sided with the unanimous Court in the companion case (remember, *Shepard's* codes *Grayned* as "following" *Mosley*), it is possible that these two justices accepted the content neutrality rule as a doctrine but were uncomfortable with the specific application in *Mosley*. Thus, their special concurrences in *Mosley* skewed the ideological coalition to be more liberal; the *Grayned* coalition is perhaps more reflective of the coalition that established the content neutrality doctrine, although *Mosley* is the opinion in which the doctrine was established.

The only significant ideological variable is the measure of absolute ideology for the opinion coalition in the *Mosley* model, but it is signed in the wrong direction. Given that *Mosley* was a liberal outcome, my expectation was that an increase in the liberalism of the coalition would lead to an increase in doctrinal utilization, not vice versa. If ideology is modeled using Martin-Quinn scores, no ideological variables attain significance.[50] In the *Mosley* model using Segal-Cover scores, authorial ideology is highly significant in the expected direction (β = 1.4143, SE = 0.473, p = .003). As ideology becomes more liberal, the topic score increases. This

[50] Martin-Quinn scores indicate increasing conservatism as the score increases (although precisely speaking, it is not a liberal-conservative scale measure). Segal-Cover scores indicate increasing liberalism as the score increases.

relationship is true for the *Grayned* model as well (β = 1.505, SE = 0.451, p = .001). In the Segal-Cover model, various justice fixed effects become significant as well.[51]

Chronological age of the doctrine is significant in the liberalism score and Segal-Cover score models, but in the wrong direction (increased age leads to increased discussion). This result does not comport with my theoretical expectations but does align with the Lowess plots. Neither of the averages for doctrinal vitality are significant in any model. In models that use the sentence proportions as the dependent variable operationalization of doctrinal usage, the only variable that attains significance is doctrinal age (in the wrong direction), so I do not report full regression results for these models.[52]

As a final test in the content neutrality area, I ran a count model of the number of sentences that contain content neutrality doctrinal language. While Poisson is appropriate for count models, goodness-of-fit testing indicates that the data is overdispersed, which makes negative binominal regression the appropriate model (see table 4).[53]

[51] Note that in the Segal-Cover *Grayned* models, the author ideological distance variable is omitted due to collinearity.
[52] Model specifications were identical to the topic score model, with the exception of using sentence proportions to calculate doctrinal vitality.
[53] The Vuong test confirmed that negative binomial regression was appropriate, rather than zero-inflated negative binomial regression.

Table 4. Negative Binominal Results for Content Neutrality Sentence Counts

Variable	***Grayned***	***Mosley***
Doctrine Age	.093 (.014)***	0.091 (.014)***
Author Ideology	-1.193 (2.968)	-.681 (2.918)
Coalition Ideology	-0.405 (.475)	-.448 (.464)
Author Ideological Distance	-3.124 (2.994)	-2.467 (2.967)
Coalition Ideological Distance	.119 (.674)	-.456 (.642)
Liberal Opinion Direction	.165 (.116)	.164 (.116)
Justice Fixed Effects	*	*
Vitality (Five-Year Average)	.117 (.067)	0.116 (.066)
Vitality (Ten-Year Average)	-.291 (.105)**	-.284 (.104)**
N	1267	1267
Wald X^2	335.51***	334.62***

* $p <= .05$ ** $p <= .01$ *** $p <= .001$. Robust standard errors in parentheses.

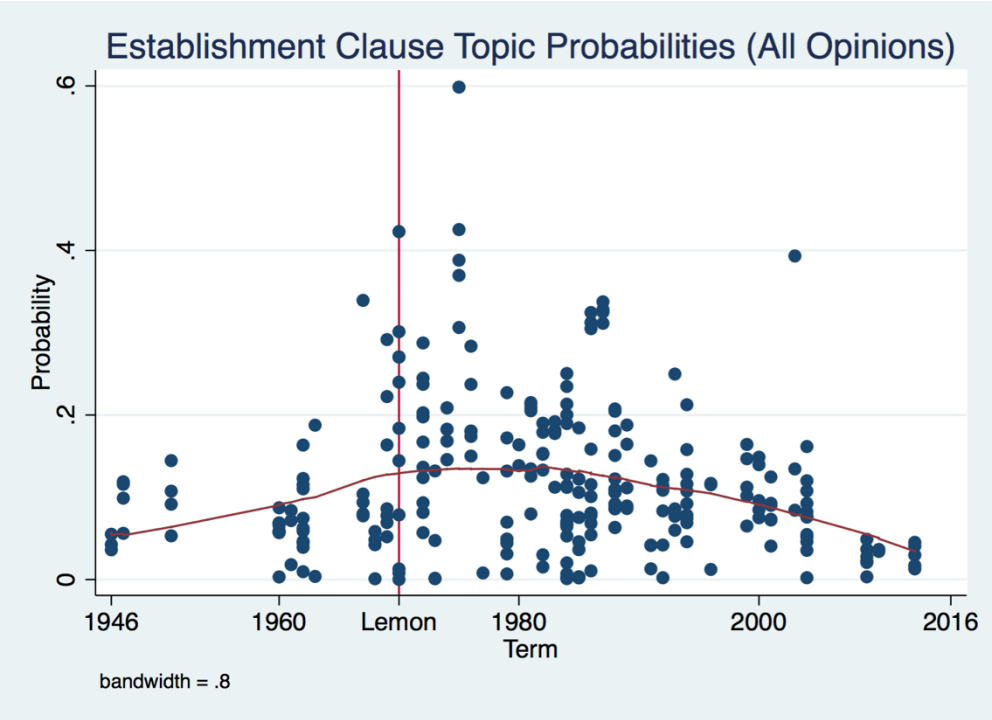
These results show some similarities to the LDA model. Doctrine age is again significant in the wrong direction. Ideology is not significant; however, the ten-year average is significant but in an unexpected direction (increased discussion over the previous ten years correlates with decreased discussion in the current opinion). The five-year average is signed in the expected direction, and approaches significance at the .05 level ($p = .079$). Under alternate ideological specifications, the ten-year average remains highly significant in the wrong direction, and the five-year average becomes significant at the .05 level. These results seem somewhat counterintuitive, in that shorter-term vitality performs as expected but longer-term vitality does not. Vitality needs further testing across other doctrinal areas to determine if this result is consistent and merits further interpretation. The only alternate ideological variable that becomes significant at the .05 level is absolute coalition ideology in the *Mosley* Martin-Quinn model. Given the lack of consistent findings in regard to ideology, there are no substantive conclusions

to draw from this result. Selected justice fixed effects are significant. In short, across all content neutrality models, there is very limited evidence to support my hypotheses (five-year vitality average), and some contradictory evidence (doctrine age).[54]

Establishment Clause – Doctrinal Utilization Models

I now turn to the analysis of doctrinal utilization in my second legal area of interest, the Establishment Clause. I used the same three dependent variable measures of doctrinal utilization that I utilized with content neutrality. I began my analysis of the Establishment Clause with a visual test of $H_{4\text{-}1}$. All three doctrinal utilization measures have similar distributions, so I only plotted LDA. These plots are seen in figure 11.

[54] In the conclusion of this chapter, I offer some thoughts as to why doctrine age may not perform as expected, but those speculations are best considered after seeing the models of the other doctrinal areas.



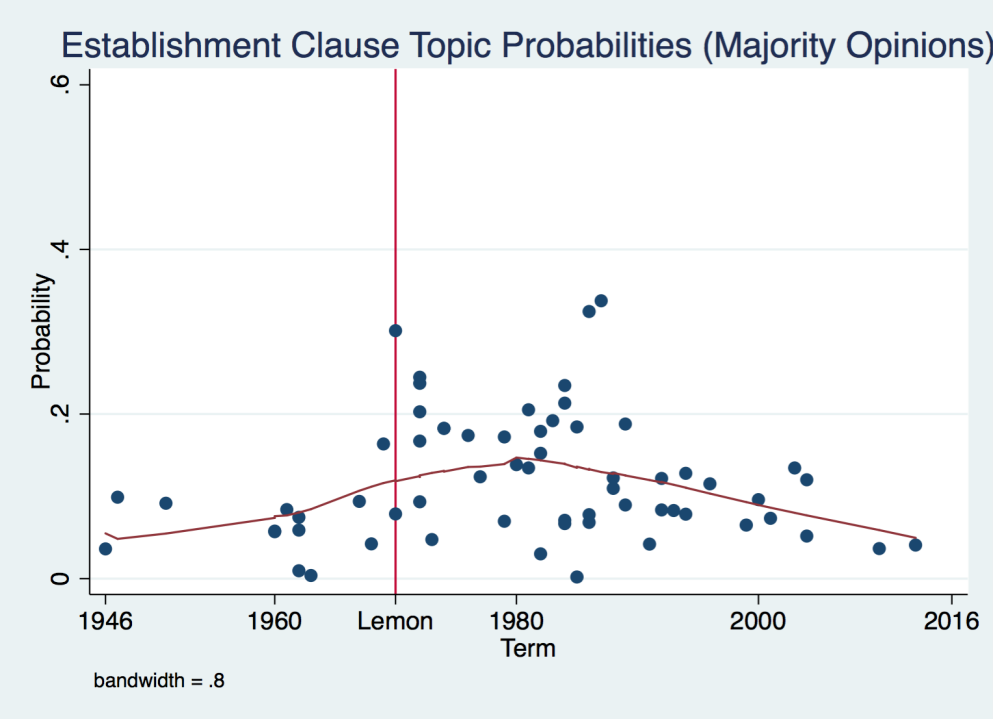


Figure 11. Establishment Clause (LDA topic probability scores)

Upon comparing these plots to the content neutrality plots, it is immediately apparent that the *Lemon* test follows my theory of doctrinal lifespan more closely than content neutrality. There is

a period prior to *Lemon* where discussion of the various doctrinal components increases. Once the *Lemon* test is fully established, discussion stabilizes for a period of time until the doctrine begins to depreciate. Thus, visual evidence indicates that $H_{4\text{-}1}$ is more promising in this area.

My regression procedures for the *Lemon* test were identical to my examination of content neutrality, with the exception of only having one reference opinion (*Lemon v. Kurtzman*).[55]

Table 5. GLM Results for Establishment Clause Topic Probabilities

Variable	***Lemon***
Doctrine Age	-.049 (.010) ***
Author Ideology	-.676 (.693)
Coalition Ideology	-.456 (.641)
Author Ideological Distance	2.092 (1.113)
Coalition Ideological Distance	.184 (.814)
Liberal Opinion Direction	.036 (.147)
Justice Fixed Effects	***
Vitality (Five-Year Average)	-5.090 (3.097)
Vitality (Ten-Year Average)	-.872 (2.385)
N	212
Wald X^2	12501.22***

* $p <= .05$ ** $p <= .01$ *** $p <= .001$. Robust standard errors in parentheses.

Doctrine age is highly significant in the expected direction. No other variables are significant, although author's ideological distance approaches significance at the .05 level ($p = .06$) in the unexpected direction. There are no differences if alternate ideological specifications are used. Unlike with content neutrality, there were some differences in the topic model and sentence proportion model; thus, I report the proportional results in table 6.[56]

[55] Fractional polynomial testing showed no significant difference in models using quadratic terms for doctrinal age and the base model.
[56] See *supra* p. 54n37 for the search rules for *Lemon* test sentences.

Table 6. GLM Results for Establishment Clause Sentence Proportions

Variable	***Lemon***
Doctrine Age	-.073 (.021)***
Author Ideology	3.839 (1.068)***
Coalition Ideology	-1.419 (.795)
Author Ideological Distance	-1.467 (1.713)
Coalition Ideological Distance	2.635 (1.370)
Liberal Opinion Direction	-.035 (.228)
Justice Fixed Effects	***
Vitality (Five-Year Average)	-1.982 (17.606)
Vitality (Ten-Year Average)	11.169 (16.363)
N	212
Wald X^2	400.45***

* $p <= .05$ ** $p <= .01$ *** $p <= .001$. Robust standard errors in parentheses.

Doctrine age remains highly significant in the expected direction. Author ideology is highly significant in the expected direction (as the author becomes more liberal, the proportion of sentences using *Lemon* test terms increases). However, coalition ideological distance is very close to significance at the .05 level ($p = .055$), but in the wrong direction. As the coalition's ideological distance from the *Lemon* coalition increases, the sentence proportion increases rather than decreases. Under the Martin-Quinn specification of ideology, these cross-cutting ideological relationships remain. In the Segal-Cover specification, coalition ideological distance is no longer significant, but author ideological distance becomes highly significant in the wrong direction. As with previous models, the ideological results are not consistent enough to draw substantively meaningful conclusions. Various justice fixed effects are significant as well. One possible modeling issue is indicated by the high standard errors on the vitality score variables, suggesting that the proportional specification of the dependent variable is inappropriate for measuring vitality scores.

Due to this standard error issue in combination with the inconsistent ideological results, I modeled sentence counts as a third way of measuring utilization. A zero-inflated model did not converge when using the liberalism or Martin-Quinn scores as a measure of ideology. The zero-inflated model did converge when using Segal-Cover scores, and post-estimation testing indicated that the zero-inflated model was to be preferred to either the negative binominal regression or Poisson models. Model diagnostics indicated that the inflation variable is the opinion type, as non-majority opinions are significantly more likely to contain no references to the *Lemon* test. This makes theoretical sense, as dissenting opinions are more likely to discuss different doctrinal considerations in an attempt to shift the conversation away from *Lemon*, and concurrences are expected to agree with the majority's doctrinal disposition and thus have less need to develop new doctrine. I report the results of the Segal-Cover ideology model in table 7 and address any differences in comparison to the negative binomial regressions using alternate ideological specifications in the discussion of results.

Table 7. Zero-Inflated Negative Binomial Results for Establishment Clause Sentence Counts

Variable	***Lemon***
Doctrine Age	-.033 (.142)*
Author Segal-Cover Ideology	-2.503 (3.396)
Coalition Segal-Cover Ideology	-1.138 (.688)
Author Segal-Cover Ideological Distance	2.347 (4.079)
Coalition Segal-Cover Ideological Distance	-.735 (1.011)
Liberal Opinion Direction	-.073 (.218)
Justice Fixed Effects	***
Vitality (Five-Year Average)	-.156 (.046)**
Vitality (Ten-Year Average)	.131 (.058)*
N	212
LR X^2	55.19***

* $p <= .05$ ** $p <= .01$ *** $p <= .001$. Robust standard errors in parentheses.

Doctrine age remains significant in the expected direction. No ideology variables are significant at the .05 level (although coalition ideology is significant at the .1 level, with a *p*-value of .09). Coalition ideology is also signed in the wrong direction. As the coalition becomes more liberal, the number of sentences devoted to the *Lemon* test decreases, rather than increases. In the Martin-Quinn specification, coalition ideology is no longer significant at the .1 level, but author ideology is significant at the .05 level (again, in the wrong direction). If any consistent ideological theme is emerging across issue areas (and it is weakly consistent at best), it is that ideology is performing in the opposite direction of my theoretical expectations. The *Lemon* test has a complex ideological history that could explain the unexpected ideology findings. It was originally approved by an 8-0 court and produced a liberal outcome, but the opinion was written by the most conservative member of the court at that time (Chief Justice Burger). The test does not appear to possess a clear ideological bent, at least based on voting behavior. Given this lack of an ideological bent, the test could be flexible enough to allow for justices on both sides of the ideological spectrum to use the test to promote their preferred outcome without abandoning the test in favor of another, more ideologically preferred doctrine.

The five-year vitality average is now highly significant, but in the wrong direction. However, ten-year vitality average is significant and signed in the correct direction. As with the content neutrality findings, the vitality averages work in opposite directions depending upon the time period. To see if these relationships continue to emerge, and to explore differences in constitutional and statutory interpretation, I turn to models involving employment discrimination.

Employment Discrimination – Doctrinal Utilization Models

Thus far, I have only found limited significant results that have accorded with my theoretical expectations (and several that have been the opposite of expectations). One potential

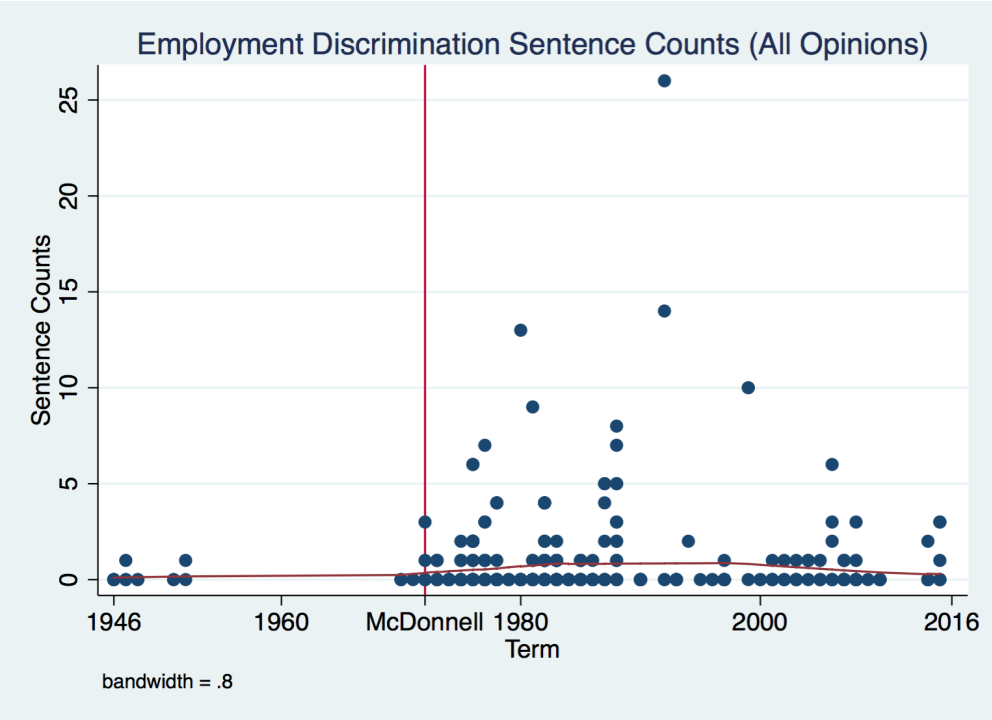
explanation for the inconsistent findings is that there is not a unified theory of doctrinal usage; that is, utilization behavior is doctrine-dependent. To explore this possibility, I turn to one example of statutory interpretation, Title VII employment discrimination law.[57] Black and Spriggs (2013) found that statutory precedents depreciate more slowly than constitutional law precedents, perhaps because the Court shows Congress more deference in statutory cases due to the ability of Congress to pass laws to overturn the Court's decision, rather than having to engage in the more difficult process of constitutional amendments.[58] I hypothesize that this deference should be apparent in the relationship between ideology and doctrinal use, in that doctrinal utilization should not exhibit sensitivity to ideological preferences in statutory interpretation. Doctrinal utilization should still exhibit sensitivity to doctrinal age and vitality, as these are not factors that should affect judicial deference.[59]

As a representative area of statutory law, I analyzed Title VII employment discrimination cases.[60] These cases are governed by the burden-shifting framework originally articulated in the 1973 case *McDonnell Douglas Corporation* v. *Green* (411 U.S. 792). In brief, *McDonnell Douglas* developed a three-step process for adjudicating employment discrimination cases. The plaintiff must first make a credible prima facie case for discrimination, at which point the burden shifts to the defendant to produce legitimate, non-discriminatory reasons for the actions taken.

[57] Title VII is a reference to Title VII of the Civil Rights Act of 1964.
[58] As with most of their findings, Black and Spriggs acknowledged that while various factors affect precedent depreciation in statistically significant ways, the effects are quite small substantively. Nonetheless, the constitutional/statutory distinction merits exploration.
[59] Future work will incorporate a more complex separation-of-powers model to analyze the relationship between Congress and the Supreme Court in statutory interpretation. The goal of this initial analysis is to detect broad differences between constitutional and statutory interpretation using the same modeling strategy.
[60] I identified the universe of cases using the Supreme Court Database's three issue areas for employment discrimination cases: sex, alien status, and all other types of discrimination.

The burden then shifts back to the plaintiff to show that discrimination could be inferred from the defendant's actions. Thus, there are a few key terms that comprise the doctrinal language of the *McDonnell Douglas* framework.[61] As with the other legal areas, I began with a visual test of $H_{4\text{-}1}$ by plotting the frequency of language with *McDonnell Douglas* as the reference opinion. For this legal area, I plotted sentence frequencies as representative of the other types of linguistic analysis, as all of the distributions are similar. The plots are seen in figure 12.

[61] Specifically, in my linguistic searches, I search for "prima" and "facie," "nondiscriminatory," "legitimate" and "reason," or some combination of these terms.



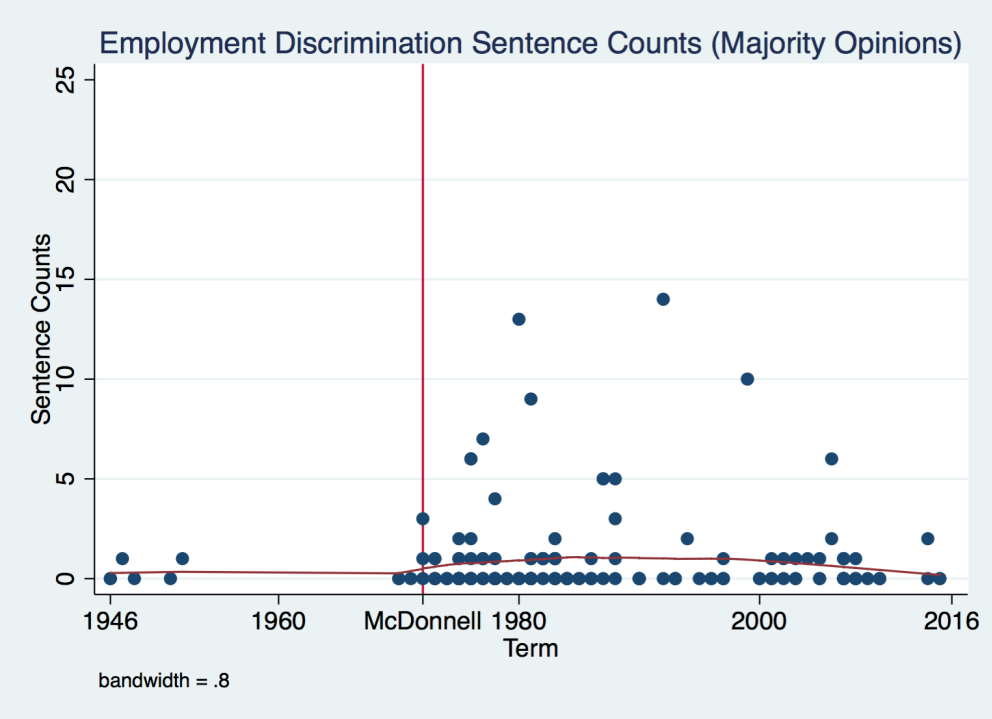


Figure 12. Employment discrimination (sentence counts)

Evidence from the visual test of $H_{4\text{-}1}$ is mixed. There does not appear to be a strong correlation between doctrinal age and doctrinal utilization, as the Lowess curve is largely flat. As

with the other legal areas, I examined the relationships using regression models. In the regression models, I again used the same three dependent variable measures of doctrinal utilization that were used in the analysis of content neutrality and the Establishment Clause, although I do not report results from the proportional model due to model specification issues (as explained in more detail below). In the topic score model using the liberalism score for ideology operationalization, there were collinearity issues with the ideological distance variables that caused those variables to be omitted, so I report results from the Martin-Quinn model in table 8.

Table 8. GLM Results for Employment Discrimination Topic Probabilities

Variable	***McDonnell Douglas***
Doctrine Age	.035 (.017)*
Author Martin-Quinn Ideology	.424 (.140)**
Coalition Martin-Quinn Ideology	-.049 (.123)
Author Martin-Quinn Ideological Distance	.170 (.129)
Coalition Martin-Quinn Ideological Distance	-.279 (.188)
Liberal Opinion Direction	.245 (.280)
Justice Fixed Effects	***
Vitality (Five-Year Average)	1.802 (.777)*
Vitality (Ten-Year Average)	.915 (1.288)
N	391
Wald X^2	361.95***

* $p <= .05$ ** $p <= .01$ *** $p <= .001$. Robust standard errors in parentheses.

As with content neutrality, doctrine age is significant, but in the wrong direction.[62] Ideology is also significant in the wrong direction (as the opinion author becomes more conservative, the topic score increases). No ideology variables are significant if Segal-Cover scores are used to

[62] Fractional polynomial testing indicates that a linear specification of age is appropriate.

model ideology.[63] The five-year average vitality score is significant in the expected direction (this is true in the Segal-Cover model as well). Similar to the Establishment Clause model, the sentence proportions model has high standard errors for the vitality score variables, indicating specification issues. Additionally, all but one justice fixed effects are highly significant in the proportional model, which is unique among the models. This is further evidence of specification issues. Thus, I do not report the proportional results and instead turn to sentence counts.

Inclusion of justice fixed effects in the sentence count model led to model misspecification, as all but one of the fixed effects variables were collinear.[64] Therefore, I dropped the justice fixed effects from the sentence count model and ran postestimation testing to determine which type of count model is the best fit. The model was overdispersed, indicating the need for a negative binominal model, and the Vuong test confirmed that a zero-inflated negative binomial was not needed. A negative binominal count model with the justice fixed effects dropped and using Martin-Quinn ideology measures is significant at the .05 level (Wald X^2= 17.25, p = 0.028). However, the only variable that is significant in the model is the ten-year doctrinal vitality average (β = -0.047, SE = 0.170, p = .006). This variable is signed in the wrong direction; as the average increases, sentence counts decrease. As with the other doctrinal areas, there are not enough statistically significant and consistent results to make strong conclusions about the hypotheses.

[63] Additionally, doctrine age is not significant in the Segal-Cover model.
[64] This is indicated by Stata's inability to calculate a likelihood ratio for the model, the inability to provide any statistics except the coefficient for the fixed effect variables, and the number of fixed effects constraints that are dropped in post-estimation parameter testing.

Conclusion

The results across all three doctrinal areas are a mixed bag: some limited support for my hypotheses, some contradictory evidence, and inconsistency that is dependent on model specifications such as which ideology measure is used. Thus, it is not possible to draw any sort of firm conclusions about doctrinal utilization on the Supreme Court level. Tentatively, it is encouraging to identify the fairly consistent existence of a relationship between doctrinal age and doctrinal utilization, even if the relationship was often inconsistent across models in terms of support for my hypothesis. This at least suggest that there is a relationship worth exploring further, and that my initial theory of utilization may have been incorrect. If I was to engage in post hoc theory building, I would note that the expected relationship appeared in Establishment Clause jurisprudence, an area where the *Lemon* test has undergone significant criticism. This level of doctrinal criticism has not been seen in the other legal areas, which possibly means that I have observed the *Lemon* test in the decline phase of its lifespan, whereas the other doctrines have not begun their decline. This would mean that my theory of doctrinal development is not necessarily wrong, just that the time slices that I have observed in this study do not demonstrate the theory. Of course, the only way to test this possibility is for more time to pass. Although there are no consistent findings about ideology, I explore an alternate way to theorize the relationship between ideology and doctrinal utilization in my study of the circuit courts in the next chapter. There is not a good post hoc explanation for the cross-cutting relationship in some of the doctrinal vitality results, so I now turn to exploration of vitality on the circuit court level to see if I can shed any additional light on the puzzle.

CHAPTER FIVE: DOCTRINAL UTILIZATION IN CIRCUIT COURT OPINIONS

Thus far, my analysis of doctrinal usage has been confined to the Supreme Court, which theoretically possesses the most flexibility in doctrinal usage due to its location at the top of the federal judicial hierarchy. In my discussion of employment discrimination jurisprudence, I briefly raised the possibility that doctrinal usage on the Supreme Court could be affected by the potential of interbranch deference to Congress. Since cases involving statutory interpretation provide a somewhat easier route for congressional intervention than cases involving constitutional interpretation, the Court is more cautious in its interpretive freedom (Black and Spriggs 2013). The question of constraint by other institutional actors is an important one, particularly in the federal judicial hierarchy.

The popular normative expectation is that vertical stare decisis acts as a significant constraint upon the lower courts in the hierarchy. Kornhauser (1995) argued that a strict respect for vertical (and horizontal) stare decisis is exactly the behavior that should be theoretically expected on the circuit courts, as it is the natural combination of adjudicating in a resource-constrained system with the goal of ensuring correctness in the law.[65] The reality is much more complex; if the Supreme Court is viewed as a principal vis-à-vis the principal-agent model, the

[65] Interestingly, Kornhauser suggests that such a system is a natural byproduct of the resource constraints of the federal hierarchy, to the point that his theory has an invisible hand flavor to it: "even if the judiciary has no substantive reason to promote certainty or uniformity, these qualities will emerge as collateral consequences of the organizational aim of getting decisions right" (1995, 1628).

lower courts have significant latitude to create moral hazards for the Supreme Court by implementing their own policy preferences.

The literature on the circuit courts has often conceptualized this problem of compliance in terms of preferred policy outcomes. A compliant circuit court votes on the merits in the way that the Supreme Court prefers; a non-compliant circuit court does not. In keeping with the theme of this project, I envision compliance somewhat differently: to what extent do circuit courts possess the freedom to elaborate their own interpretations of legal doctrine? Theoretically, if circuit courts possess the ability to shirk the outcome wishes of the Supreme Court, they should also possess the ability to go their own way in the crafting of legal doctrine. To explore this question of doctrinal compliance, I begin with a review of previous findings regarding the ideological, institutional, and hierarchical factors that influence decision-making in the circuit courts. I link these findings to my theory of doctrinal vitality and argue that many of the same factors that promote compliance or non-compliance in voting behavior should be apparent in opinion writing, although in different ways. To test my expectations for opinion-writing, I construct an original dataset of Establishment Clause cases across all federal circuits and apply my text analysis methodologies to analyze doctrinal utilization behavior. I find that both institutional and hierarchical factors play a significant role in shaping the utilization of doctrine in opinion writing in the circuit courts.

Ideology, Hierarchy, and the Circuit Courts

The central premise of the attitudinal model of judicial decision-making is that justices vote for the case outcome that best aligns with their personal policy preferences. Segal and Spaeth (2002) noted a key limitation of this model: it only applies to Supreme Court decision-making, due to the unconstrained position of the Supreme Court as a court of last resort.

Reviewing the literature, they found that "though the evidence is mixed, examination of appellate court decisions in several different issue areas shows little overtly noncompliant behavior" (Segal and Spaeth 2002, 96). In principal-agent terms, the appellate courts were found to be generally well-behaved agents. Their observation that the evidence is mixed is key, as the literature continues to find more ideological freedom among appellate judges than would be desired in an ideal principal-agent model. George put it succinctly: "Most courts of appeals judges vote their attitudes" (1998, 1965). However, George also noted that a minority of judges seem to be constrained from voting their attitudes both by institutional and hierarchical factors.

Klein (2002) offered an explanation for the mixed evidence. While circuit court judges comply with Supreme Court precedents in areas of settled law, they do not attempt to anticipate the Court's decisions in areas of developing law. He found that circuit judges are seriously committed to compliance with the law as an institutional norm. One judge went so far as to say that "as an intermediate court, we're very constricted. Often our personal inclinations are at odds with the result that must flow from the cases" (Klein 2002, 23). If this statement is to be believed, then it makes sense that more ideological freedom exists in newly developing legal areas because the judges are not as constricted.[66] Analyzing search and seizure cases, Klein found little evidence that judges attempted to anticipate how the Supreme Court would vote in a case, either in terms of voting evidence or the language of opinions. Noting the limitations of analyzing a narrow area of law over a brief period of time, he still concluded that judges do not fear reversal to the extent that an ideal principal-agent model would anticipate. Klein and Hume

[66] One is reminded of the Spaeth-Mendelson exchange where Spaeth argued that he studied votes because votes could be trusted and words could not (Mendelson 1963, 1964, 1966; Spaeth 1965).

(2003) expanded the timeframe of analysis and arrived at a similar conclusion. The fear of reversal is not an adequate explanation for compliance on the circuit courts.

In an earlier study of search-and-seizure cases, Songer, Segal, and Cameron (1994) found that both ideology and the Supreme Court's preferences were highly significant predictors of voting behavior on the circuit courts. However, due to the low rate of reversals in their sample, the authors were unable to draw any conclusions as to if congruence with Supreme Court preferences was actually the result of monitoring by the Court, or if it was simply a result of behaving according to the expectations of institutional norms that had been internalized by circuit judges. Although their study did not shed explanatory light on potential reasons for compliance with the Supreme Court, they did identify a key theoretical reason for the ability of justices to vote their ideological preferences: the difficulties that the Supreme Court faces in monitoring circuit courts. The authors noted that "because the Supreme Court for the most part learn only of circuit decisions that are appealed to them, and can act only on those that are appealed, most decisions of the courts of appeals will escape consideration by the Supreme Court" (Songer, Segal, and Cameron 1994, 675). The extremely low grant rate of certiorari petitions adds to this difficulty. Once a case is brought to the Court's attention, it is still highly unlikely to be reviewed. In this way, the Supreme Court exercises so-called "fire alarm oversight" (McCubbins and Schwartz 1984), which allows for more ideological freedom as the circuit courts (agents) can shirk the Supreme Court's (principal) wishes without a realistic fear of consequences.

Cross (2005) observed that this resource constraint faced by the Supreme Court means that the chances of reversal by the Court are so low for any given circuit court decision that any positive political theory insights about behaving rationally by responding to the preferences of

the Court to prevent reversal are theoretically interesting but practically irrelevant. He found that to the extent that judges were responsive to the Court, they were actually responsive to the ideological direction of the previous decade of precedents, indicating a respect for the law rather than any strategic anticipation of the future behavior of the Court. However, this respect for the law only partly explained appellate judge voting behavior; again, the low threat of oversight meant that ideology also influenced circuit judges. Put differently, ideology clearly matters to circuit judges, but does not tell the whole story (Cross 2003). However, the evidence does suggest that ideology tells an increasingly significant part of the story as one moves up the judicial hierarchy. Zorn and Bowie (2010) labeled this finding the "hierarchy postulate," and they found evidence that appellate judges are more ideologically free than district court judges. This finding makes intuitive sense, as district court judges are subject to mandatory oversight because of the automatic right of review of district court cases.

Given the way that the heightened potential for oversight may dampen the possibility of ideological voting, some studies have identified the potential for "whistleblowing" effects on divided panels (Sunstein, Schkade, and Ellman 2004). Cross and Tiller (1998) found that among Republican panels on the circuit courts, the effect of ideology was "dramatically" lower in 2-1 panels where a potential whistleblower existed to draw attention to the panel's disobedience of current legal doctrine.[67] Kastellec (2011a) suggested that this whistleblowing effect is highly sensitive to the ideological alignment between the whistleblowing judge and the current Supreme Court. Panel judges are more likely to serve as compliant agents when they know that the whistleblowing judge is ideologically congruent with the current Court and is thus more likely to

[67] Interestingly, they noted that an alternative explanation for this effect that accounts for the low likelihood of reversal by the Supreme Court is that whistleblowers can actually make a judge aware of their non-compliance and that awareness will be enough to change their behavior.

send a signal that is heard. This whistleblowing effect does not prevent all ideologically-based shirking but does significantly increase the ability of the Supreme Court to enact its policy preferences (Kastellec 2007). Interestingly, the possibility of hierarchical control of ideological voting through the presence of a whistleblower is a relatively recent phenomenon on the circuit courts (Kastellec 2011b).

The mixed findings of the literature in regard to the ability of the judicial hierarchy to dampen the effect of circuit judge ideology on voting behavior also has implications for the process of opinion writing. Circuit judges justify their decisions by appealing to ideologically preferred precedents (Niblett and Yoon 2015). This finding suggests that ideological considerations shape more than just the vote on the merits; they also shape the way that circuit judges attempt to craft the content of the law moving forward. Lax (2007) described the process of arriving at a legal rule on appellate courts in terms of developing a collegial median rule that is essentially an ideological compromise between the judges on the panel. The rule may not be the preferred rule of any single judge; however, it is close enough to the preferred rule to allow a majority of judges to sign on to the rule. Hinkle (2017) found evidence that the presence of a whistleblower on a panel can change not only voting behavior but the ideological content of opinions. Specifically, the presence of a whistleblower affects the decision to take an interpretive posture toward a particular precedent, which suggests that the ideological preference for precedents identified by Niblett and Yoon (2015) is tempered by whistleblowers.

On the other hand, Black and Owens (2013) found that judges who are ideologically distant from the majority coalition are much more likely to attempt to bargain for changes in the content of the opinion when compared to ideologically congruent judges but are less likely to be successful in their bargaining. Judges clearly take the ideological direction of the legal content of

an opinion seriously, even if they are not always able to shape the opinion to produce a legal rule with which they can be comfortable. Ideology also matters for the decision to dissent from the majority opinion. The decision to write separately is also influenced by collegiality norms and case salience, but ideological disagreement remains a central driving factor (Hettinger, Lindquist, and Martinek 2003). Hettinger, Lindquist, and Martinek found that circuit judges who choose to dissent are driven more by their ideological distance from the majority coalition than they are by any "strategic" attempts to change the outcome of the case by forcing a "circuit intervention" (2004, 123). Choi and Gulati (2008) found that strategic attempts to shape the ideological content of opinions come not in the decision to dissent, but in the process of what they call "trading votes for reasoning." In short, they found evidence that judges will cast votes in favor of a less preferred outcome in order to join the majority's opinion and thus gain the ability to push the doctrinal components of the opinion as close to their preference as they can.

The general takeaway from the literature on ideology and hierarchy in the circuit courts is that judges are significantly influenced by their ideological concerns, both in their voting behavior and in the way in which they try to shape opinion content. However, these ideological preferences are tempered by hierarchical concerns, both in terms of respect for the prior law laid down by the Supreme Court and in consideration of the possibility that they could be reversed if they are too non-compliant with the preferences of the Court. Judges are quite realistic about how the resource constraints of the federal judicial hierarchy impact the likelihood of Supreme Court auditing; thus, they feel relatively free to vote their ideological preferences unless circumstances like the presence of a potential whistleblower significantly increase the chances of review. Judges also feel more freedom to act ideologically in legal areas where the Court has not clearly spoken. Circuit judges do not appear to attempt to anticipate how the Court will vote in

an uncertain area; instead, they take advantage of the uncertainty to craft the law in their preferred ideological direction.

Given this interaction between ideology and hierarchical constraints, I expect doctrinal discussion in circuit opinions to behave differently than expected on the Supreme Court level. On the Supreme Court level, I presented some evidence that an increase in ideological distance from a doctrine-establishing justice or coalition leads to a decrease in discussion of the doctrine. This is because of the incentive structure that Supreme Court justices face in shaping the law. Justices who want a particular legal doctrine to remain vital in that area of jurisprudence can maintain the doctrine's vitality through extensive discussion. On the other hand, justices who do not want a doctrine to remain a vital part of jurisprudence attempt to steer doctrinal development down a different course by changing the subject and turning attention to a new doctrine. On the circuit court level, the incentive structure is reversed. Since judges are aware that non-compliance could lead to increased monitoring and the possibility of reversal if someone was to blow the whistle, circuit judges will feel an increased need to dedicate significant discussion to a doctrine with which they disagree as a way of explaining their non-compliance.[68] Circuit judges who are in ideological alignment with the doctrine can be briefer in their explanations, as they can simply cite to the Supreme Court's discussion. In terms of a formal hypothesis:

$H_{5\text{-}1}$: On the circuit court level, as the ideological distance between an opinion author or panel increases relative to the Supreme Court author or panel that established the doctrine, discussion of the doctrine will increase.

[68] Westerland et al. (2010) noted that this fear of reversal decreases if the contemporary Court is ideologically distant from the Court that established a precedent, as the contemporary Court will be less likely to uphold the precedent.

On the Supreme Court level, I provided evidence that discussion of a doctrine naturally declines over time for two primary reasons: societal changes making the doctrine increasingly irrelevant and less need to discuss the doctrine as the meaning and scope becomes crystallized. Increased discussion of a doctrine is necessary when the doctrine is originally established to define the doctrine, but that necessity decreases over time. By extension, the more a doctrine is discussed on the Supreme Court level (thus indicating increased uncertainty as to the doctrine's meaning and application), the more freedom circuit courts should possess to place their own ideological spin on the meaning of the doctrine.

Corley and Wedeking (2014) made a similar argument using measures of linguistic certainty based on dictionaries in the Linguistic Inquiry and Word Count program. They found that as the Supreme Court became more certain with its language in discussing a precedent, the circuit courts were more likely to comply with the Court's decision. Uncertainty has also been operationalized as a function of coalition size, as circuits are more likely to comply with majority opinions rather than plurality opinions (Corley 2009). Smith and Todd (2015) noted that the Supreme Court will use rules rather than standards as a tool to increase compliance by circuit courts, because rules are more concrete and thus more likely to force compliance (see also Baker and Kim 2012). Connecting my expectations about increased discussion representing more ambiguity with the ideological expectations of $H_{5\text{-}1}$, I hypothesize a formal relationship between Supreme Court doctrinal vitality and discussion on the circuit courts:

$H_{5\text{-}2}$: As doctrinal vitality increases on the Supreme Court level, discussion of the doctrine will increase on the circuit court level.

This hypothesis draws some inspiration from the formal model of Carrubba and Clark (2012), who argued that circuit judges who want to produce legal rules that are ideologically

disparate from the Supreme Court's preferences will write higher quality opinions. Higher quality opinions are more compelling in terms of logic and are thus harder to overturn. My theory is using amount of discussion as a rough proxy for quality of argument, in that more discussion of a doctrine indicates a more carefully argued opinion. Importantly, while I theorize a relationship between doctrinal age and discussion on the Supreme Court level, I do not theorize the same on the appellate court level. Since doctrinal utilization on the circuit court is dependent upon doctrinal utilization on the Supreme Court, utilization on the circuit courts is not expected to experience a depreciation effect that independently responds to time.

Ideology and Unpublished Opinions

One of the unique features of the circuit courts that impacts the effect of judge ideology is the high frequency of unpublished opinions. Unpublished opinions (also known as memorandum dispositions) are not authorized for publication in the *Federal Reporter*, and thus many circuits have adopted rules against treating these opinions as precedential. Kozinski and Reinhardt, two judges who served on the Ninth Circuit, once offered a very simple explanation for the Ninth Circuit's rule against citing unpublished opinions: "If memodispos could be cited as precedent, conscientious judges would have to pay much closer attention to their precise wording" (2000, 44, see also Robel 1989; Wasby 2001). Circuit judges see unpublished opinions as speaking to the disposition of the particular case at hand, rather than articulating a general legal rule that can apply across multiple cases. They serve as a time-saving function for circuits that are swamped by large caseloads.[69]

[69] Kozinski and Reinhardt noted that in 1999 alone, the Ninth Circuit decided 4,500 cases on the merits, publishing 700 opinions and 3,800 memoranda (2000, 44).

An obvious ideological expectation that follows from this understanding of unpublished opinions is that judge ideology should be less relevant in the disposition of cases that are unpublished. If circuit judges do not treat unpublished opinions as precedential documents that produce legal rules, then there is no reason for judges to be concerned with how the opinions may shape the ideological direction of the law going forward. In a study of litigation involving the U.S. Forest Service, Keele et al. (2009) found this exact relationship: ideology affected judge behavior in published opinions but not in unpublished opinions. Law (2005) identified evidence of strategic behavioral shifts in asylum cases decided in the Ninth Circuit, in that Democrat panels were more likely to grant asylum in cases that were published than in cases that were unpublished, presumably to push the precedential law in a pro-asylum direction.[70] Given that circuit judges are generally unconcerned with the development of legal rules in unpublished opinions, the logical expectation is that unpublished opinions will contain less doctrinal discussion.[71] Put formally:

$H_{5\text{-}3}$: Unpublished opinions will contain less doctrinal discussion.

Strategic Auditing and Institutional Factors

In my discussion of circuit judge ideological effects, I noted that the limited oversight resources that the Supreme Court possesses allows for significant ideological freedom among circuit judges. However, the Court is able to use strategic auditing methods to identify when more oversight of a circuit may be necessary. Several informational shortcuts have been identified. Scott (2006) found that the Court is more likely to exercise its oversight powers over

[70] Interestingly, no strategic effects were identified among Republican panels.
[71] Note that this relationship is not simply because unpublished opinions are invariably shorter. The causal arrow flows in the other direction – unpublished opinions are shorter because of the lack of doctrinal discussion.

an ideologically distant circuit, an intuitive expectation that I accounted for in my ideology hypothesis (see also Cameron, Segal, and Songer 2000; Lindquist, Haire, and Songer 2007). Lindquist and her co-authors also found that prior interactions between the circuit and the Supreme Court partially explained future monitoring behavior, in that circuits that had been reversed more frequently in the past were more likely to be reviewed by the Court in the future (Lindquist, Haire, and Songer 2007). This is another intuitive finding, as non-compliant agents are more likely to exhibit continued non-compliance relative to compliant agents.

Another strategic signal of the need for Supreme Court oversight can be sent using en banc review. The Supreme Court is more likely to review cases that are decided en banc. George and Solimine (2001) suggested that one possible explanation for this finding is that en banc cases serve as a signal as to the legal significance of a decision. The Court is more likely to grant certiorari because the fact that the circuit took the time to rehear a case en banc indicates that the issue is salient, complex, and worthy of further consideration. Circuit judges who desire to blow the whistle on non-compliant behavior by a panel can use dissents as a way to signal the need for en banc review. Beim, Hirsch, and Kastellec (2016) found that judges who blow the whistle on panels that they are close to the panel ideologically carry more credibility because they are sending a signal that runs counter to their personal preferences. Given this counterintuitive finding, judges in this position seem to possess a significant desire to get the law right, since en banc opinions are more susceptible to review by the Supreme Court. By sending a signal to the full circuit that review is needed, the dissenting judge may be communicating a signal upwards through the judicial hierarchy. Consequently, I hypothesize that en banc opinions will contain increased discussion of legal doctrine for salience reasons and to lessen the probability of review and reversal by the Supreme Court. Assuming that circuit courts want to avoid increased

oversight by the Court, the full circuit needs to dedicate increased discussion to the justification of their decision, more so than a panel that is highly unlikely to be reviewed en banc, all else being equal. These interactions between institutional considerations and hierarchical review can be formally expressed with two hypotheses:

$H_{5\text{-}4}$: As a circuit is reversed more frequently by the Supreme Court, the amount of discussion in their opinions that is dedicated to doctrine will increase.

$H_{5\text{-}5}$: En banc opinions will contain more doctrinal discussion than panel opinions.

While my institutional discussion to this point has focused on the location of the circuit courts in the federal judicial hierarchy, intracircuit and intercircuit relationships affect circuit behavior as well. All circuits are bound by precedent produced within their own circuit, but they are not formally bound by precedents produced by other circuits. Intracircuit compliance with precedent is not perfect but generally strong. Using *Shepard's Citations*, Lee (2003) noted that Sixth Circuit judges rarely treated intracircuit precedents negatively, and when they did, it was usually the weak negative treatment of distinguishing. Negative treatment of intracircuit precedents is more likely to occur in published opinions rather than unpublished, providing further support to the idea that ideology is more important in published opinions (Lee 2005). Since circuit judges see themselves as generally compelled to follow intracircuit precedents, I hypothesize that increased discussion of a legal doctrine within a circuit will lead to increased discussion of the doctrine in future opinions issued by the circuit. Put formally:

$H_{5\text{-}6}$: An increase in intracircuit doctrinal vitality will increase the amount of discussion of the doctrine in future circuit opinions.

Although they are not formally bound to do so, Klein (2002) noted that circuit judges admit to drawing on the precedents produced by other circuits as a way of promoting consistency

in the law. Hume (2009) discovered that various features of administrative law opinions increased the likelihood that the precedents would be favorably cited by other circuits. One of these features is the amount of block quotations that an opinion contains, working on the assumption that increased block quotation of material indicates more thorough grounding in supporting evidence. Similarly, I hypothesize that increased discussion of a doctrine in a circuit opinion indicates more careful discussion of that doctrine, and thus increases the likelihood that other circuits will feel compelled to discuss the doctrine in their opinions.

$H_{5\text{-}7}$: An increase in intercircuit doctrinal vitality will increase the amount of discussion of the doctrine in future circuit opinions.

Establishment Clause – The Circuit Courts Opinions Dataset

To test these expectations for doctrinal utilization in the circuit courts, I constructed a dataset of Establishment Clause opinions issued across all circuits. In addition to being one of the legal areas analyzed in the Supreme Court chapter, previous research on First Amendment religion decisions in the circuit courts has revealed significant ideological effects upon judge behavior. Judges appointed by Republican presidents are more likely to vote in a pro-religion manner and vice-versa for judges appointed by Democrats (Heise and Sisk 2012; Sisk and Heise 2012), although other work has suggested that this effect may be at least partly the result of religious beliefs (Sisk, Heise, and Morriss 2004). On the other hand, there is evidence that circuit courts are compliant with the legal doctrine promulgated in Establishment Clause jurisprudence, independent of the ideological preferences of circuit judges (Luse et al. 2009). Thus, the Establishment Clause provides a great test for the impact of ideological and hierarchical effects upon doctrinal utilization on the circuit courts.

The doctrine of interest is the *Lemon* test, so I began my analysis with opinions issued in 1972 or later (as *Lemon* was issued by the Supreme Court in 1971). To identify candidate opinions to analyze, I conducted a LexisNexis search of all circuit court opinions that are tagged with the keywords "Establishment Clause" or "Establishment of Religion." I ensured that the opinions contained a substantive discussion of Establishment Clause doctrine by eliminating any opinions where the circuit did not reach an Establishment Clause merits issue (e.g., opinions where the substantive issue was standing only or where the circuit court remanded to the district court to decide the Establishment Clause conflict as an issue of first impression). My dataset contains panel and en banc opinions in addition to unpublished opinions (although the LexisNexis database only contains unpublished opinions issued in 2001 or later). This collection process yields 692 opinions from 444 cases (all opinion types included) across 12 circuits (the Federal Circuit's jurisdiction does not include Establishment Clause cases).[72] For descriptive purposes, table 9 delineates the dataset by circuit, opinion type, publication status, and en banc review status.

[72] The dataset contains opinions issued through December 2017.

Table 9. Establishment Clause Opinions, Circuit Courts – 1972-2017

Circuit	Majority	Dissent	Concurrence	Special	En Banc	Unpublished
1	12	6	4	1	0	0
2	36	11	1	3	0	1
3	31	13	7	0	2	5
4	38	19	7	4	15	0
5	30	12	3	1	3	1
6	57	27	15	5	14	5
7	42	18	6	0	7	2
8	34	17	4	0	4	0
9	70	26	8	14	14	8
10	33	17	2	1	11	1
11	28	8	5	2	8	1
DC	9	1	1	3	0	0
All	420	175	63	34	78	24

Note: All categories follow the coding rules of the Supreme Court Database. "Special" refers to special concurrences, which can refer to concurring in the judgment only or concurring in part and dissenting in part.

As in the Supreme Court chapter, I modeled doctrinal utilization three ways: LDA topic scores, *Lemon* test sentence counts, and *Lemon* test sentence proportions relative to the opinion length. The dependent variables were calculated using the text analysis methodologies outlined in the previous chapters. The LDA topic score was calculated using a model that only includes the circuit court opinions. Twenty topics is the appropriate model, and topic clusters are included in the appendix. Although I do not expect any significant relationship between the age of the *Lemon* test and amount of discussion in circuit opinions, I provide Lowess-smoothed plots of the relationship between the opinion year and the three dependent variables (for all opinion types – see figures 13, 14, and 15).

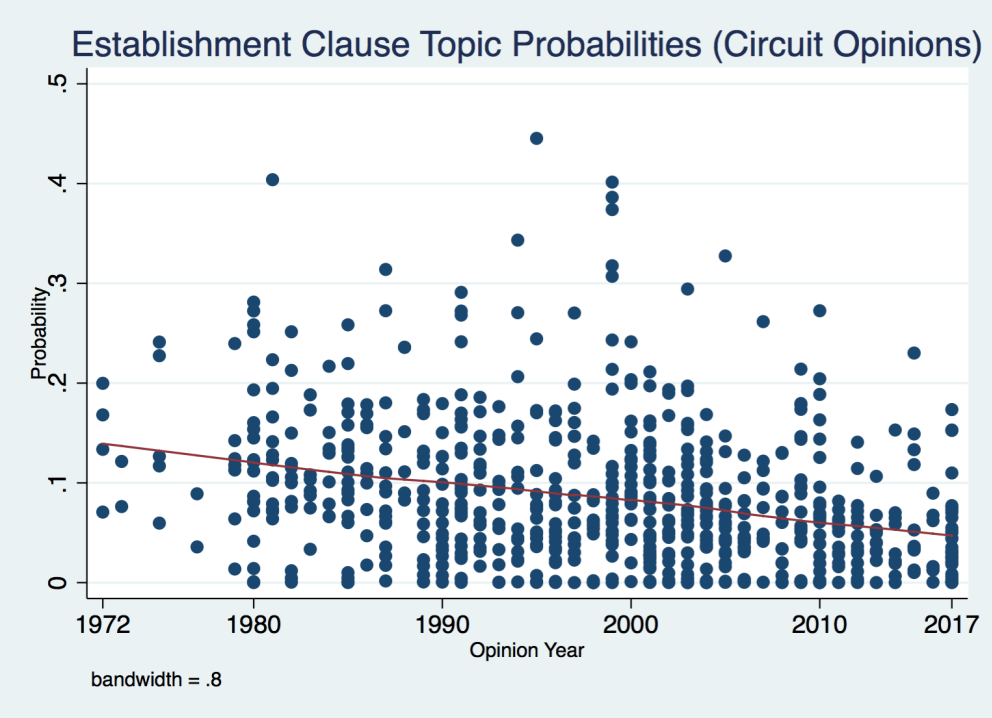


Figure 13. Establishment Clause – circuit opinions (LDA topic probability scores)

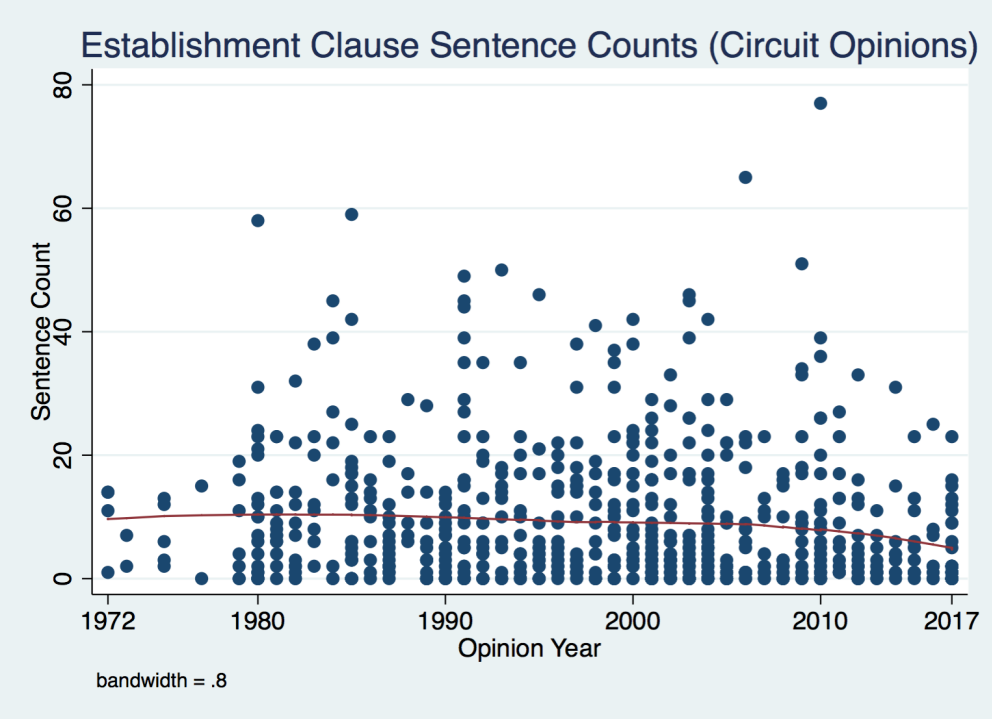


Figure 14. Establishment Clause - circuit opinions (sentence counts)

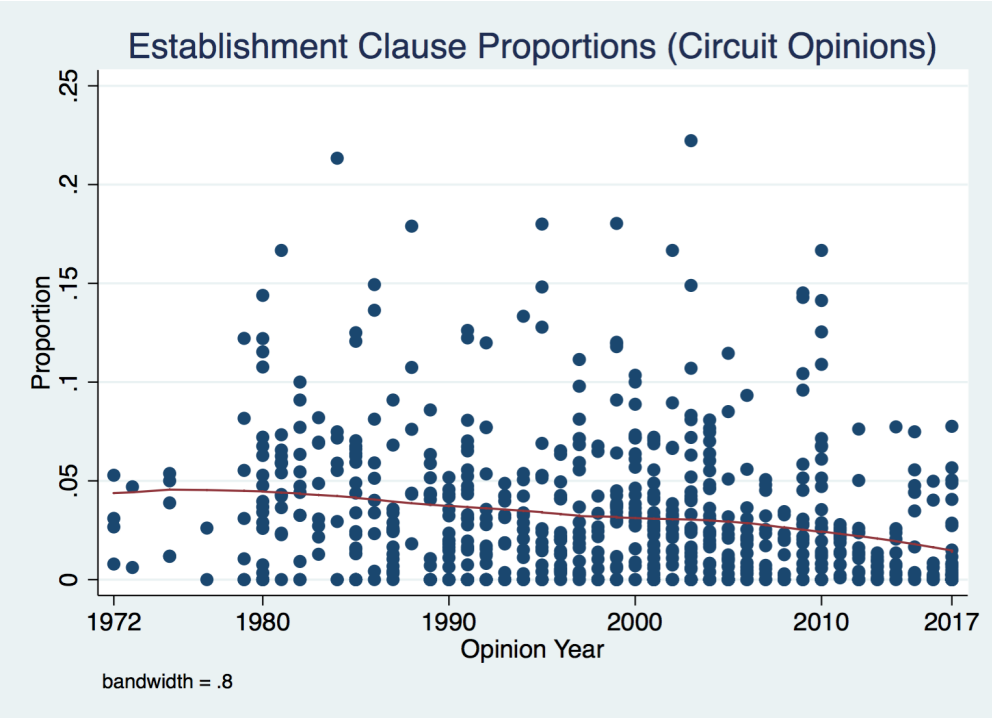


Figure 15. Establishment Clause - circuit opinions (sentence proportions)

The plots indicate a slight decline in doctrinal discussion over time, although no consistently strong relationship is evident. The decline is most pronounced when doctrinal utilization is modeled using LDA topic scores. The distribution of the dependent variable is similar across all three plots, indicating similarity in the linguistic concept being measured.

Doctrinal Utilization – Model Variables

My dependent variable for each model was the measure of doctrinal utilization (topic score, sentence count, and sentence proportion).[73] For each opinion, I included a control variable that measured the number of years that have passed since *Lemon* was issued by the Supreme

[73] For this analysis, I assume that doctrinal utilization can be measured in the same way on the Supreme Court level and the circuit court level. It is theoretically possible (although unexpected) that there are some systemic differences in how Supreme Court justices and circuit court judges discuss doctrine. In future research, I will validate the doctrinal utilization measures on the circuit court level as I did on the Supreme Court level (with manual opinion analysis) to explore this possibility.

Court. To test my expectations regarding en banc and unpublished opinions, I included a dummy variable that indicated if an opinion was either one of those types. Ideological effects were operationalized in two ways: a liberalism score within Establishment Clause cases in the dataset, and the Judicial Common Space score for each judge (Epstein et al. 2007).[74] Using these two scores, I included four ideology variables in each model. Two of the variables measured the absolute ideology of the author of the opinion and the coalition that signed on to the opinion (for opinions signed only by the author, these are the same). The other two ideology variables measured the relative distance between the opinion author and Chief Justice Burger when he authored *Lemon*, and the coalition that signed on to the opinion and the *Lemon* majority coalition. As with the Supreme Court models, I included a variable that controlled for the ideological direction of the opinion (coded using Supreme Court Database rules).[75]

I included three different measures of doctrinal vitality in each model (as in the previous chapters, the operationalization of doctrinal vitality matched the way that the dependent variable was measured). To control for hierarchical effects, I included a five-year and ten-year average of doctrinal vitality scores from Supreme Court opinions.[76] These scores were calculated beginning at the Supreme Court term prior to the year in which the circuit court opinion was issued to ensure an appropriate lag to affect circuit court behavior. I included five-year and ten-year

[74] On rare occasion, a retired Supreme Court justice sits by designation on a circuit panel. In these instances, the liberalism score was constructed using only their votes on the circuit level, rather than imputing their score from the Supreme Court level. I used this strategy to account for the possibility that the justice may behave differently when they are sitting at a different level of the hierarchy. Not every judge in the dataset has a Judicial Common Space score, due to how the scores are calculated (e.g., when a justice from the Court of International Trade sits by designation). I drop these opinions from the model, explaining the difference in sample size between models that use different operationalizations of ideology. The JCS scores are current as of the October 8, 2016 update.

[75] Conservative opinions (1) are pro-accommodationist; liberal opinions (2) are pro-separationist.

[76] These averages are taken from the Supreme Court models in chapter four.

vitality averages for majority opinions issued across all circuits and a separate score only for majority opinions issued within the same circuit. To control for circuit effects not captured by doctrinal vitality measures, I included a fixed effects dummy for each circuit. As another control for the influence of justices and judges outside of the circuit, I included a dummy variable to identify when an opinion was written by a justice or judge sitting by designation.

To measure reversal rates, I used the Supreme Court Database to calculate a running rate of the percentage of times that the circuit has been reversed prior to the year in which the opinion was issued.[77] This reversal rate was calculated based only on cases where the Supreme Court chose to grant certiorari. If the Court did not grant certiorari, then I did not include that case in the calculation. This has the effect of overstating the true reversal rates of circuits, because the Court grants certiorari in cases in which it is likely to overturn the lower court, and a denial of cert has the effect of leaving a lower court decision in place (and thus, not reversing the circuit). However, since I employed this measurement strategy across all circuits, the overstating of the reversal rate did not bias my results. I separately modeled the reversal rate only in Establishment Clause cases to capture issue-specific non-compliance.[78] All reversal rates were calculated beginning with Supreme Court cases decided in the 1946 term.

Doctrinal Utilization – Model Results

I begin with reporting the results of the LDA topic score model in table 10.[79]

---

[77] I consider reversals to be any action taken by the Supreme Court that does not fully uphold the case as decided by the circuit court.

[78] The following circuits did not have an Establishment Clause case heard before the Supreme Court in the 1946-2016 terms: First, Fourth, Seventh, Tenth, Eleventh, DC, and Federal Circuit. Thus, in the model where I included Establishment Clause reversal rates, I only included opinions from circuits that have been reversed.

[79] The 1989 Seventh Circuit per curiam opinion in *Mather* v. *Mundelein* (864 F.2d 1291) was included in the LDA analysis for building the topic models but was not included in the regression analyses. While it was signed as a per curiam opinion, two members of the panel filed separate

Table 10. GLM Results for Establishment Clause Probabilities (Circuit Courts)

Variable	**Liberalism Score**	**JCS Score**
Doctrine Age	-.049 (.018)**	-.049 (.018)**
Unpublished	-.138 (.228)	-.201 (0.245)
En Banc	-.441 (.123)**	-.443 (.129)**
Designation Author	.010 (.163)	.079 (.174)
Ideology Direction	-.079 (.113)	-.085 (.073)
Author Ideology	.063 (.172)	-1.721 (1.984)
Coalition Ideology	-.064 (.144)	.288 (.292)
Author Ideological Distance	.111 (.245)	-1.448 (2.012)
Coalition Ideological Distance	.016 (.240)	-.479 (.430)
Supreme Court Vitality (Five-Year Average)	-2.530 (2.457)	-2.288 (2.472)
Supreme Court Vitality (Ten-Year Average)	.787 (2.904)	.131 (3.039)
All Circuits Vitality (Five-Year Average)	-15.171 (5.415)**	-16.702 (5.497)**
All Circuits Vitality (Ten-Year Average)	4.509 (9.444)	6.008 (9.259)
Circuit Vitality (Five-Year Average)	1.070 (.888)	1.086 (.890)
Circuit Vitality (Ten-Year Average)	-1.259 (1.096)	-1.333 (1.083)
Circuit Reversal Rate	1.673 (.715)**	1.644 (.737)*
Circuit Fixed Effects	**	**
N	691	680
Wald X^2	151.28***	153.36***

* p <= .05 ** p <= .01 *** p <= .001. Robust standard errors in parentheses.

Although not theoretically expected, there is a highly significant negative relationship between doctrine age and topic probability. There is also a highly significant relationship between en banc opinions and topic probability, but in the wrong direction. En banc opinions

concurring opinions, and one member of the panel filed a separate dissenting opinion. These separate opinions make it difficult to know what the ideological alignment of the per curiam opinion was, so I omitted the per curiam opinion and included the separate opinions in the regressions. This explains the one opinion difference in the number of opinions issued in 1972 or later and the *N* in the regression models.

have lower topic probability scores. The five-year average vitality score for all circuit opinions is highly significant, but also in the wrong direction. As vitality increases across all circuits, the topic probability in future opinions decreases. However, the coefficient and standard error for this variable is questionable, so further modeling is needed. The circuit reversal rate is highly significant in the expected direction. As a circuit's reversal rate increases, the topic probability increases as well. Fixed effects for opinions issued by the Fourth and Fifth circuit were significant in a positive direction, indicating that these circuits have higher topic scores compared to other circuits. Contrary to expectations, no ideology variables were significant.[80]

In my discussion of the Supreme Court results, I noted some modeling issues that resulted from how the topic scores and sentence proportions were measured. Similar issues are apparent on the circuit level, especially the inflated robust standard errors for the vitality score variables. Running a model with *Lemon* sentence proportions produced the same inflated standard errors and essentially the same substantive results. Doctrine age, en banc opinions, and reversal rates remain significant in the same direction as in the topic probability model. Some circuit fixed effects are significant (Second and Fourth positively, DC circuit negatively, Sixth and Seventh positively at the .1 level). No ideology variables are significant, and the results remain substantively unchanged if JCS scores are used to measure ideology.

The inflated robust standard errors are corrected in the sentence count models on the circuit level, although the sentence count models do present a zero-inflation problem in the dependent variable. Using the opinion vote as the inflator as I did on the Supreme Court level, post-estimation testing of the zero-inflated negative binomial regression indicated that it is to be preferred over either a Poisson or negative binomial regression. However, the opinion vote is not

[80] There are no substantive changes to the results if the circuit fixed effects variable is removed.

actually inflating the zero counts, as it does in the Supreme Court models (the *p*-values for the inflate variable are all non-significant). Additionally, unlike on the Supreme Court level, there is no good theoretical reason to expect the opinion type to inflate the zeroes, as all opinion types are likely to contain doctrinal discussion due to the potential need to signal for en banc or Supreme Court review. There are no other variables in the model that are theoretically expected to inflate the zeros, so I report the results of a negative binomial regression in table 11 and note any substantive differences with the zero-inflated model in discussion of the results.

Table 11. Negative Binomial Results for Sentence Counts (Circuit Courts*)*

Variable	**Liberalism Score**	**JCS Score**
Doctrine Age	-.005 (.007)	-.003 (.007)
Unpublished	-1.088 (.207)***	-1.179 (.208)***
En Banc	-.785 (.153)***	-.639 (.162)***
Designation Author	-.232 (.179)	-.210 (.191)
Ideology Direction	.930 (.139)	-.220 (.094)*
Author Ideology	-.353 (.232)	2.347 (2.844)
Coalition Ideology	-.174 (.193)	.254 (.445)
Author Ideological Distance	.647 (.330)*	2.585 (2.891)
Coalition Ideological Distance	.080 (.318)	-2.462 (.716)**
Supreme Court Vitality (Five-Year Average)	.006 (.017)	-.003 (.016)
Supreme Court Vitality (Ten-Year Average)	.033 (.038)	.035 (.039)
All Circuits Vitality (Five-Year Average)	-.022 (.027)	-.034 (.028)**
All Circuits Vitality (Ten-Year Average)	.022 (.046)	.015 (.047)
Circuit Vitality (Five-Year Average)	.022 (.007)**	.018 (.007)*
Circuit Vitality (Ten-Year Average)	-.032 (.010)**	-.025 (.010)**
Circuit Reversal Rate	1.411 (1.677)	.971 (1.507)
Circuit Fixed Effects	**	**
N	691	680
Wald X^2	115.84***	153.36***

* p <= .05 ** p <= .01 *** p <= .001. Robust standard errors in parentheses.

For both ideology models, unpublished opinions have lower doctrinal sentence counts, in line with my expectations. En banc opinions have higher sentence counts, contradicting my initial expectations. There are mixed results for the ideology variables. In the liberalism score model, the author's ideological distance from Burger is significant in the expected direction. As the author becomes more unlike Burger ideologically, the author devotes more discussion to the *Lemon* test. However, this relationship does not hold in the JCS model. In the JCS model, the ideological distance from the *Lemon* coalition is significant, but in the unexpected direction. As distance increases, the amount of discussion devoted to the *Lemon* test decreases. It is worth noting that the JCS ideology scores may not be as trustworthy as the liberalism scores. The JCS scores are determined across all issues areas, not just Establishment Clause cases. The JCS scores also mask nuance in the ideological differences between judges, because the coding scheme assigns the same scores to many judges.[81] Thus, it is encouraging to see some support for my ideology hypothesis in the liberalism model, but the questions surrounding the JCS measure mean that firm conclusions about ideology cannot be drawn.

The strongest consistent statistical finding in both models regards vitality scores on the intracircuit level. Within a circuit, increased discussion over the previous five case years is correlated with increased discussion in the current opinion. Paradoxically, increased discussion over the past ten case years is correlated with decreased discussion in the current opinion. Mathematically, this result is counterintuitive. While increased discussion in the immediately preceding case years is expected to correlate with increased discussion in the current opinion, there is not an immediately apparent reason why a long-term increase in discussion would lead to decreased current discussion. Perhaps the long-term discussion creates a "discussion fatigue"

[81] Note the somewhat higher standard errors for the absolute JCS variables.

effect. Under this theory, a short-term increase in discussion points to the need to speak to a live controversy about the doctrine, whereas a long-term increase indicates that the circuit has been in the midst of the doctrinal discussion for a while, and there is not much left to add. However, this interpretation is speculative and cuts against the grain of my general hypotheses regarding doctrinal vitality. There is one significant intercircuit vitality finding in the JCS model, but it is not supported by the other intercircuit variables and is signed in the unexpected direction.

The only interesting difference in the zero-inflated models using the opinion vote as the inflator is that in the liberalism score model, as an author gets more liberal, the sentence count for the *Lemon* test decreases. This finding aligns with my ideology hypothesis but is not supported by the other models and can thus only be taken as tentative support for the hypothesis. The vitality scores remain paradoxically significant in both the liberalism score and JCS model. As with the negative binomial regression model, various circuit fixed effects are significant, but these circuits change depending upon the model and there are no strong theoretical reasons to expect significance from any particular circuit.

It is worth briefly noting a model that includes a measure of the circuit rate of reversals in Establishment Clause cases. Since several circuits have never had an Establishment Clause case heard before the Supreme Court, the sample in this model is significantly smaller (N=327). I do not report full regression results for this model because there are no substantive changes to variables already analyzed in previous models, but the reversal rate in previous Establishment Clause cases is a significant predictor of increased doctrinal discussion in both ideology models (at the .05 level). In the JCS model, the coefficient for the reversal rate is 0.914 and the standard error is 0.412, and in the liberalism score model the coefficient is 0.946 and the standard error is 0.453. These results align with my hypothetical expectations. As a circuit is reversed more

frequently in Establishment Clause cases, it devotes more discussion to the *Lemon* test in future opinions, perhaps to provide more justification for its reasoning to guard against future oversight.

Conclusion

Given the amount of methodological ground covered, it is helpful to briefly review my findings relative to my hypotheses. For the ideology hypothesis ($H_{5\text{-}1}$), there was mixed support depending on the model, and some findings contradicted the hypothesis. There was no support for the Supreme Court doctrinal vitality hypothesis ($H_{5\text{-}2}$), potentially counterintuitive support for intracircuit vitality ($H_{5\text{-}6}$), and some evidence to contradict the intercircuit vitality hypothesis ($H_{5\text{-}7}$). In the sentence count models, there was strong support for my expectations regarding unpublished opinions ($H_{5\text{-}3}$), but en banc opinions tended to perform contrary to expectations across all models ($H_{5\text{-}5}$). It is perhaps the case that my expectations for en banc opinions were initially backwards. Being as en banc opinions are more likely to be reviewed by the Supreme Court, the full circuit may want to express more deference to the Court by simply citing to Court opinions, rather than including more discussion of the doctrine. This would accord with my general expectation that increased discussion on the circuit courts happens when a judge wants to disagree with the Court rather than agree with the Court. This post hoc explanation is tentative, however, and merits further exploration. Finally, while doctrinal discussion is not sensitive to overall circuit reversal rates, it is highly sensitive to reversal rates in Establishment Clause cases for both ideology models. This provides support for the reversal hypothesis ($H_{5\text{-}4}$).

CHAPTER SIX: CONCLUSION

This dissertation project began with two key assumptions: legal doctrine matters, and the language used to express legal doctrine is substantively important. While previous literature had explored the importance of legal precedents through citation patterns and interpretation, I set out to explore the actual doctrinal content of legal opinions. Previous studies used automated text analysis tools to examine linguistic features such as tone and shared language, but the overarching goal of this project was a different one: could I isolate the doctrinal content of legal opinions and identify various factors that influence the decision to utilize this language? At the conclusion of this dissertation, it is helpful to review my findings and examine the extent to which this project contributes to the literature on legal doctrine from a theoretical, methodological, and substantive perspective.

In brief, I believe the project challenges the literature to carefully consider the meaning of legal doctrine relative to the concept of precedent and demonstrates methodological issues that can arise when those concepts are conflated. Methodologically, this dissertation demonstrates the strengths and weaknesses inherent in various text analysis tools when used to measure the utilization of doctrine. While the substantive takeaways are mixed in terms of identifying consistent predictors of doctrinal usage, my findings do point to important relationships to consider as the question of doctrinal utilization continues to be explored. In this conclusion, I highlight the key findings from each chapter and discuss how those findings point to future questions that merit exploration. I also address issues related to the scope of my results and what obstacles to generalization exist as the work moves forward.

In terms of theoretical contributions, I argued for conceptual clarity in understanding legal doctrine as one component of a precedent. A single precedent consists of three broad parts: the case facts, the legal doctrine that is the decision rule, and the case outcome. As demonstrated in chapter two, previous studies that depend upon *Shepard's Citations* as a measure of the interpretation of precedent often conflate the concepts of precedent and doctrine, and thus are not adequate tools for understanding the utilization of legal doctrine specifically.[82] I demonstrated the consequences of conflating precedent and doctrine through case studies of how *Shepard's* treats *Lemon* v. *Kurtzman* to show how the interpretation of *Lemon* is not correlative with usage of the *Lemon* test.

I did not critique *Shepard's* as a tool for measuring the interpretation of a particular precedent; on that score, it has been demonstrated to be a valid and reliable tool.[83] I instead argued that measuring the utilization of a legal doctrine is a fundamentally separate question from measuring the interpretation of a precedent. Since the utilization of doctrine is a substantively interesting question in and of itself for the future trajectory of the law, my methodological goal was to develop tools that captured the utilization of legal doctrine. The substance of legal doctrine is expressed in the language of judicial opinions; thus, I was able to leverage text analysis tools that have been utilized in the judicial politics literature.

Chapter three was dedicated to methodological issues surrounding the measurement of legal doctrine. I evaluated two tools that are used to examine document similarity, plagiarism

---

[82] I am indebted to scholars such as Deborah Beim, Thomas Hansford, Benjamin Kassow, Kelly Rader, Joseph Smith, and Maureen Stobb for challenging me to more precisely define the theoretical aims of this project. Put differently, they pushed me to answer the "so what?" question in sharper terms.

[83] I am especially grateful to Matthew Hitt for pointing to the need to clarify that my critique was not aimed at *Shepard's* as a tool in and of itself. Rather, I am asking a question that *Shepard's* is not designed to answer.

detection and cosine similarity. Both of these tools can be useful measures of legal doctrine if an opinion draws a large amount of doctrinal language from a precedential opinion but are less effective for analyzing the use of doctrinal language on a sentence level. To accomplish this task, I combined the unsupervised topic modeling method of called Latent Dirichlet Allocation and the quasi-supervised method of sentence counts that were based upon my specifications of doctrinal language. While I demonstrated the validity of these tools for capturing the utilization of doctrine, it is important to note some methodological limitations that should be considered for future research.

One limitation is the use of topic modeling as a measure of the doctrinal proportion of an opinion, rather than its usual usage as a document classification technique. The topics that LDA modeling identified in each doctrinal area contained words associated with the doctrines of interest but were not isolated to those words in particular.[84] Thus, the topic scores in substantive models are best seen as rough estimates of the prevalence of doctrinal language in an opinion. Given how an LDA model is calculated, the estimates are also relative to the prevalence of other topics in the opinion. As a result, it is not substantively useful to think in terms of fine-grained increases and decreases of the doctrinal topic score, but rather to think in terms of the broader concept of doctrinal language being more prevalent in an opinion relative to other opinions.

In some ways, this note of caution is applicable for the quasi-supervised sentence models as well. My broad assumption that justified the use of sentence count models is that higher sentence counts clearly translate into more discussion of a doctrine. Since the sentence count methodology allows me to control the specific language being counted (unlike in LDA modeling,

[84] As seen in the appendix, the doctrinal words were not always the primary words that comprised the topic.

where the supervision is much more limited), sentence count models also increase the precision with which I am able to model the use of doctrinal language. However, there are theoretical questions about the use of language that are left unexplored. For instance, is there really a substantive difference in doctrinal utilization in an opinion that contains three sentences discussing the doctrine versus in an opinion that contains six sentences? What if an opinion only contains one sentence that discusses the doctrine, but that one sentence contains the only doctrinal content in the entire opinion, and is thus extremely important for the logic for the opinion? While I somewhat address this second issue with the sentence proportions measures in my statistical models, that solution is only partial and suffers from the reality that some opinions are going to be longer than others for a myriad of reasons that are not reducible to quantitative measures. Even though this project has made great methodological strides toward the measurement of doctrinal language, the reality remains that language is qualitative and statistics are quantitative.[85]

Measuring language quantitatively will always present reliability and validity challenges that do not exist in the measurement of something like judicial voting behavior. Votes are relatively easy to code; language is not. As Segal and Spaeth noted, the English language presents a large combination of possibilities for expressing ideas. Given this flexibility, there is always the open question of how far the influence of doctrinal language can rightly be felt. How many synonyms need to be measured? Is there a limit to the extent that synonyms exhibit

[85] Another example of a limiting assumption of my project is my assumption that increased utilization of language is necessarily positive (at least on the Supreme Court level). It could of course be that justices talk more about a doctrine to limit its utilization going forward, rather than promote the doctrine. This seems like a question for sentiment analysis and could be explored in further research; however, sentiment analysis has met with limited success in the study of judicial politics because of fundamental differences in how language is used in the law versus among the general public.

conscious influence by the original articulation of the legal doctrine? Does linguistic influence need to be conscious to count as influence? How much intentionality exists in the decision to use one combination of language over another? Do the differences in the combinations matter? Drawing from Spaeth's concerns about the ability to trust votes and not words, how much stock can be put in language as a reliable indicator of judicial thought patterns? Is discussion of a doctrine a reliable proxy for the influence of that doctrine in judicial thinking? These questions do not have easy answers, and I have not attempted to assuage any and all validity and reliability concerns that could possibly be raised. Rather, my goal in this project has been to raise important theoretical and methodological questions related to the development of legal doctrine, and to demonstrate the reliability and validity of some possible methodological solutions to this very complex problem.

Since there will always be at least some loss of information in the attempt to measure qualitative language with quantitative statistics, my various findings in the substantive chapters on doctrinal utilization on the Supreme Court and circuit courts should be interpreted as broadly indicative trends, rather than as fine-grained analyses that have substantive predictive value. In other words, although the statistical models claim to do so, the substantive import of my findings is not that a one-year increase in doctrinal age leads to a one sentence in doctrinal discussion (as a hypothetical example). That level of precision is simply impossible when measuring language and does not make sense substantively in terms of how language is used. What does it mean to interpret my findings in terms of broader trends? On the Supreme Court level, there are relationships between doctrinal age and doctrinal language in an opinion (although that relationship varies depending on the doctrinal area). On the circuit court level, unpublished opinions contain less doctrinal discussion on average. These two examples are meant to

demonstrate the broader way in which to interpret my findings. It is a story about broad trends, not fine-grained counts.

Given these limitations of the measurement of language, the fact that I was able to both measure language and demonstrate statistically significant (and substantively meaningful) relationships between factors such as age, ideology, opinion characteristics, and doctrinal utilization does highlight significant promise for this area of research moving forward. It is also important to note that the legal areas that I chose to study are somewhat unrepresentative of the larger universe of legal doctrine. As Hansford and Spriggs have noted in previous work, the majority of precedents are cited once or not at all. In other words, many legal doctrines are likely not nearly as prevalent or as contested in opinions as the doctrines that I have studied in this project. Thus, it may not be possible to develop a so-called grand unified theory of doctrinal language usage; the usage likely varies from legal area to legal area (as has been seen in this study). Intuitively, this usage varies along the dimension of contestability as well, in that more contested legal areas (such as the ones studied here) likely demonstrate different patterns of linguistic usage than less-contested areas. These are the types of issues to explore as this research is expanded beyond the doctrinal areas studied in this dissertation, but they should always be explored with an appreciation of the limitations of generalizability in mind.

On the scale of generalizability, this project falls somewhere between the careful case studies that allow for understanding the linguistic nuance of legal doctrine, and the broad studies of precedent citation and interpretation. It should be fairly apparent that translating some of this methodology to an entire universe of Supreme Court cases, for instance, may be impossible, or at the least, extremely burdensome in terms of research resource constraints. However, to the extent that this project advances our knowledge of doctrinal utilization in legally important areas of

doctrinal development, then this project makes a needed contribution to the literature. To the extent that this project challenges us to think in terms of conceptual clarity as to what matters in the development of the law (the doctrine/precedent distinction), then this project makes a vital contribution to the literature. To the extent that this project makes advances in measuring a qualitative concept quantitatively and demonstrates that it is possible to understand doctrinal development in ways beyond case studies, then this project makes an important contribution to the literature. The law matters; thus, understanding the way in which judges shape the law through the language of opinions matters. That has been the mission of this study; it is up to the reader to determine the extent to which that mission has been accomplished.

REFERENCES

Aguilar v. Felton. 1985. 473 U.S. 402.

Baker, Scott, and Pauline T. Kim. 2012. "A Dynamic Model of Doctrinal Choice." *Journal of Legal Analysis* 4 (2):329-63.

Bartels, Brandon L., and Andrew J. O'Geen. 2015. "The Nature of Legal Change on the U.S. Supreme Court: Jurisprudential Regimes Theory and Its Alternatives." *American Journal of Political Science* 59 (4):880-95.

Beim, Deborah, Alexander V. Hirsch, and Jonathan P. Kastellec. 2016. "Signaling and Counter-Signaling in the Judicial Hierarchy: An Empirical Analysis of En Banc Review." *American Journal of Political Science* 60 (2):490-508.

Benesh, Sara C., and Harold J. Spaeth. 2007. "The Constraint of Law: A Study of Supreme Court Dissensus." *American Politics Research* 35 (5):755-68.

Bird, Steven, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python*: O'Reilly Media, Inc.

Black, Ryan C., and Ryan J. Owens. 2013. "Bargaining and Legal Development in the United States Courts of Appeals." *American Politics Research* 41 (6):1071-101.

Black, Ryan C., and James F. Spriggs. 2013. "The Citation and Depreciation of U.S. Supreme Court Precedent." *Journal of Empirical Legal Studies* 10 (2):325-58.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993-1022.

Board of Education of Central School District No. 1 v. Allen. 1968. 392 U.S, 236.

Bueno de Mesquita, Ethan, and Matthew Stephenson. 2002. "Informative Precedent and Intrajudicial Communication." *The American Political Science Review* 96 (4):755-66.

Cameron, Charles M., Jeffrey A. Segal, and Donald Songer. 2000. "Strategic Auditing in a Political Hierarchy: An Informational Model of the Supreme Court's Certiorari Decisions." *The American Political Science Review* 94 (1):101-16.

Cardozo, Benjamin N. 1921. *The Nature of the Judicial Process*. New Haven, CT: Yale University Press.

Carrubba, Cliff, Barry Friedman, Andrew D. Martin, and Georg Vanberg. 2012. "Who Controls the Content of Supreme Court Opinions?" *American Journal of Political Science* 56 (2):400-12.

Carrubba, Clifford J., and Tom S. Clark. 2012. "Rule Creation in a Political Hierarchy." *The American Political Science Review* 106 (3):622-43.

Chang, Jonathan, Jordan Boyd-Graber, Sean Garrish, Chong Wang, and David M. Blei. 2009. "Reading Tea Leaves: How Humans Intepret Topic Models." *Advances in Neural Information Processing Systems*:288-96.

Choi, Stephen J., and G. Mitu Gulati. 2008. "Trading Votes for Reasoning: Covering in Judicial Opinions." *Southern California Law Review* 81:735-80.

Clark v. Community for Creative Non-Violence. 1982. 468 U.S. 288.

Cohen, Linda R. 1994. "Politics and the Courts: A Comment on McNollgast Symposium on Positive Political Theory and Law." *Southern California Law Review* 68:1685-90.

Collins, Paul M., Pamela C. Corley, and Jesse Hamner. 2014. "Me Too: An Investigation of Repetition in U.S. Supreme Court Amicus Curiae Briefs." *Judicature* 97 (5):228-34.

———. 2015. "The Influence of Amicus Curiae Briefs on U.S. Supreme Court Opinion Content." *Law & Society Review* 49 (4):917-44.

Committee for Public Education and Religious Liberty v. Regan. 1980. 444 U.S. 646.

Conkle, Daniel O. 1993. "Lemon Lives." *Case Western Reserve Law Review* 43:865-82.

Corley, Pamela C. 2008. "The Supreme Court and Opinion Content: The Influence of Parties' Briefs." *Political Research Quarterly* 61 (3):468-78.

———. 2009. "Uncertain Precedent: Circuit Court Responses to Supreme Court Plurality Opinions." *American Politics Research* 37 (1):30-49.

Corley, Pamela C., Paul M. Collins, and Bryan Calvin. 2011. "Lower Court Influence on U.S. Supreme Court Opinion Content." *The Journal of Politics* 73 (1):31-44.

Corley, Pamela C., and Justin Wedeking. 2014. "The (Dis)Advantage of Certainty: The Importance of Certainty in Language." *Law & Society Review* 48 (1):35-62.

County of Allegheny v. American Civil Liberties Union, Greater Pittsburgh Chapter. 1989. 492 U.S. 573.

Cross, Frank B. 2003. "Decisionmaking in the U.S. Circuit Courts of Appeals." *California Law Review* 91 (6):1457-515.

———. 2005. "Appellate Court Adherence to Precedent." *Journal of Empirical Legal Studies* 2 (2):369-405.

Cross, Frank B., and Emerson H. Tiller. 1998. "Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals." *The Yale Law Journal* 107 (7):2155-76.

Cutter v. Wilkinson. 2005. 544 U.S. 709.

D'Elia-Kueper, Justine, and Jeffrey A. Segal. 2015. "The Evolving View of the Law and Judicial Decision-Making." In *Emerging Trends in the Social and Behavioral Sciences*: John Wiley & Sons, Inc.

Epstein, Lee, and Jack Knight. 1998. *The Choices Justices Make*. Washington, D.C. : CQ Press.

Epstein, Lee, Andrew D. Martin, Jeffrey A. Segal, and Chad Westerland. 2007. "The Judicial Common Space." *Journal of Law, Economics, & Organization* 23 (2):303-25.

Evans, Michael, Wayne McIntosh, Jimmy Lin, and Cynthia Cates. 2007. "Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research." *Journal of Empirical Legal Studies* 4 (4):1007-39.

Everson v. Board of Education of the Township of Ewing. 1947. 330 U.S. 1.

Fix, Michael P., Justin T. Kingsland, and Matthew D. Montgomery. 2017. "The Complexities of State Court Compliance with U.S. Supreme Court Precedent." *Justice System Journal* 38:149-63.

Fowler, James H., and Sangick Jeon. 2008. "The authority of Supreme Court precedent." *Social Networks* 30 (1):16-30.

Fowler, James H., Timothy R. Johnson, James F. Spriggs, Sangick Jeon, and Paul J. Wahlbeck. 2007. "Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court." *Political Analysis* 15 (3):324-46.

George, Tracey E. 1998. "Developing a Positive Theory of Decisionmaking on U.S. Courts of Appeals." *Ohio State Law Journal* 58:1635-96.

George, Tracey E., and Lee Epstein. 1992. "On the Nature of Supreme Court Decision Making." *The American Political Science Review* 86 (2):323-37.

George, Tracey E., and Michael E. Solimine. 2001. "Supreme Court Monitoring of the United States Courts of Appeals En Banc." *Supreme Court Economic Review* 9:171-204.

Gillman, Howard. 2001. "What's Law Got to Do with It? Judicial Behavioralists Test the "Legal Model" of Judicial Decision Making." *Law & Social Inquiry* 26 (2):465-504.

Grayned v. City of Rockford. 1972. 408 U.S. 104.

Griffiths, Thomas L., and Mark Steyvers. 2004. "Finding scientific topics." *Proceedings of the National Academy of Sciences* 101 (suppl 1):5228-35.

Grimmer, Justin. 2010. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." *Political Analysis* 18 (1):1-35.

Hansford, Thomas G., and James F. Spriggs, II. 2006. *The Politics of Precedent on the U.S. Supreme Court*. Princeton, N.J.: Princeton University Press.

Hansford, Thomas G., James F. Spriggs, and Anthony A. Stenger. 2013. "The Information Dynamics of Vertical Stare Decisis." *The Journal of Politics* 75 (4):894-906.

Harris v. McRae. 1980. 448 U.S. 297.

Heise, Michael, and Gregory C. Sisk. 2012. "Religion, Schools, and Judicial Decision Making: An Empirical Perspective." *The University of Chicago Law Review* 79 (1):185-212.

Hernandez v. Commissioner of Internal Revenue. 1989. 490 U.S. 680.

Hettinger, Virginia A. , Stefanie A. Lindquist, and Wendy L. Martinek. 2003. "Separate Opinion Writing On The United States Courts Of Appeals." *American Politics Research* 31 (3):215-50.

Hettinger, Virginia A., Stefanie A. Lindquist, and Wendy L. Martinek. 2004. "Comparing Attitudinal and Strategic Accounts of Dissenting Behavior on the U.S. Courts of Appeals." *American Journal of Political Science* 48 (1):123-37.

Hinkle, Rachael K. 2015. "Legal Constraint in the US Courts of Appeals." *The Journal of Politics* 77 (3):721-35.

———. 2017. "Panel Effects and Opinion Crafting in the US Courts of Appeals." *Journal of Law and Courts* 5 (2):313-36.

Hinkle, Rachael K., Andrew D. Martin, Jonathan David Shaub, and Emerson H. Tiller. 2012. "A Positive Theory and Empirical Analysis of Strategic Word Choice in District Court Opinions." *Journal of Legal Analysis* 4:407-44.

Hume, Robert J. 2009. "The Impact of Judicial Opinion Language on the Transmission of Federal Circuit Court Precedents." *Law & Society Review* 43 (1):127-50.

Hunt v. McNair. 1971. 403 U.S. 945.

Hunt v. McNair. 1973. 413 U.S. 734.

Jacobi, Tonja, and Emerson H. Tiller. 2007. "Legal Doctrine and Political Control." *Journal of Law, Economics, & Organization* 23 (2):326-45.

Kassow, Benjamin, Donald R. Songer, and Michael P. Fix. 2012. "The Influence of Precedent on State Supreme Courts." *Political Research Quarterly* 65 (2):372-84.

Kastellec, Jonathan P. 2007. "Panel Composition and Judicial Compliance on the US Courts of Appeals." *Journal of Law, Economics, & Organization* 23 (2):421-41.

———. 2011a. "Hierarchical and Collegial Politics on the U.S. Courts of Appeals." *The Journal of Politics* 73 (2):345-61.

———. 2011b. "Panel Composition and Voting on the U.S. Courts of Appeals over Time." *Political Research Quarterly* 64 (2):377-91.

Keele, Denise M., Robert W. Malmsheimer, Donald W. Floyd, and Lianjun Zhang. 2009. "An Analysis of Ideological Effects in Published Versus Unpublished Judicial Opinions." *Journal of Empirical Legal Studies* 6 (1):213-39.

Klein, David E. 2002. *Making Law in the United States Courts of Appeals*. New York, NY: Cambridge University Press.

Klein, David E., and Robert J. Hume. 2003. "Fear of Reversal as an Explanation of Lower Court Compliance." *Law & Society Review* 37 (3):579-606.

Knight, Jack, and Lee Epstein. 1996. "The Norm of Stare Decisis." *American Journal of Political Science* 40 (4):1018-35.

Kornhauser, Lewis A. 1995. "Adjudication by a Resource-Constrained Team: Hierarchy and Precedent in a Judicial System Symposium on Positive Political Theory and Law." *Southern California Law Review* 68:1605-30.

Kozinski, Alex, and Stephen Reinhardt. 2000. "Please Don't Cite This!" *California Lawyer* 43.

Kritzer, Herbert M., and Mark J. Richards. 2003. "Jurisprudential Regimes and Supreme Court Decisionmaking: The Lemon Regime and Establishment Clause Cases." *Law & Society Review* 37 (4):827-40.

———. 2005. "The Influence of Law in the Supreme Court's Search-and-Seizure Jurisprudence." *American Politics Research* 33 (1):33-55.

———. 2010. "Taking and Testing Jurisprudential Regimes Seriously: A Response to Lax and Rader." *The Journal of Politics* 72 (2):285-8.

Lamb's Chapel v. Center Moriches Union Free School District. 1993. 508 U.S. 384.

Landes, William M., and Richard A. Posner. 1976. "Legal Precedent: A Theoretical and Empirical Analysis." *The Journal of Law & Economics* 19 (2):249-307.

Larson v. Valente. 1982. 456 U.S. 228.

Lauderdale, Benjamin E., and Tom S. Clark. 2012. "The Supreme Court's Many Median Justices." *American Political Science Review* 106 (4):847-66.

———. 2014. "Scaling Politically Meaningful Dimensions Using Texts and Votes." *American Journal of Political Science* 58 (3):754-71.

Law, David S. 2005. "Strategic Judicial Lawmaking: Ideology, Publication, and Asylum Law in the Ninth Circuit." *University of Cincinnati Law Review* 73:817-66.

Lax, Jeffrey R. 2007. "Constructing Legal Rules on Appellate Courts." *The American Political Science Review* 101 (3):591-604.

———. 2011. "The New Judicial Politics of Legal Doctrine." *Annual Review of Political Science* 14 (1):131-57.

———. 2012. "Political Constraints on Legal Doctrine: How Hierarchy Shapes the Law." *The Journal of Politics* 74 (3):765-81.

Lax, Jeffrey R., and Kelly T. Rader. 2010a. "Legal Constraints on Supreme Court Decision Making: Do Jurisprudential Regimes Exist?" *The Journal of Politics* 72 (2):273-84.

———. 2010b. "The Three Prongs of a Jurisprudential Regimes Test: A Response to Kritzer and Richards." *The Journal of Politics* 72 (2):289-91.

lda developers. 2017. *lda: Topic modeling with latent Dirichlet Allocation* 2014 [cited January 2 2017]. Available from http://pythonhosted.org/lda/.

Lee, Emergy G., III. 2003. "Horizontal Stare Decisis on the U.S. Court of Appeals for the Sixth Circuit." *Kentucky Law Journal* 92:767-92.

Lee, Emery G., III. 2005. "Precedent Direction and Compliance: Horizontal Stare Decisis on the U.S. Court of Appeals for the Sixth Circuit." *Seton Hall Circuit Review* 1:5-26.

Lee v. Weisman. 1992. 505 U.S. 577.

Lemon v. Kurtzman. 1971. 403 U.S. 602.

Levi, Edward H. 1948. "An Introduction to Legal Reasoning." *The University of Chicago Law Review* 15 (3):501-74.

Levitt v. Committee for Public Education & Religious Liberty. 1973. 413 U.S. 472.

Lindquist, Stefanie A., and Frank B. Cross. 2005. "Empirically Testing Dworkin's Chain Novel Theory: Studying the Path of Precedent." *New York University Law Review* 80:1156-206.

Lindquist, Stefanie A., Susan B. Haire, and Donald R. Songer. 2007. "Supreme Court Auditing of the US Courts of Appeals: An Organizational Perspective." *Journal of Public Administration Research and Theory: J-PART* 17 (4):607-24.

Luse, Jennifer K., Geoffrey McGovern, Wendy L. Martinek, and Sara C. Benesh. 2009. "“Such Inferior Courts . . .”: Compliance by Circuits with Jurisprudential Regimes." *American Politics Research* 37 (1):75-106.

Lynch v. Donnelly. 1984. 465 U.S. 668.

Maltzman, Forrest, James F. Spriggs, and Paul J. Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game*. New York: Cambridge University Press.

Marsh v. Chambers. 1983. 463 U.S. 783.

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis* 10 (2):134-53.

McCubbins, Mathew D., and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science* 28 (1):165-79.

McDonnell Douglas Corporation v. Green 1973. 411 U.S. 792.

McGuire, Kevin T., Georg Vanberg, Charles E. Smith, and Gregory A. Caldeira. 2009. "Measuring Policy Content on the U.S. Supreme Court." *The Journal of Politics* 71 (4):1305-21.

McNollgast. 1994. "Politics and the Courts: A Positive Theory of Judicial Doctrine and the Rule of Law Symposium on Positive Political Theory and Law." *Southern California Law Review* 68:1631-84.

Meek v. Pettinger. 1975. 421 U.S. 349.

Mendelson, Wallace. 1963. "The Neo-Behavioral Approach to the Judicial Process: A Critique." *The American Political Science Review* 57 (3):593-603.

———. 1964. "The Untroubled World of Jurimetrics." *The Journal of Politics* 26 (4):914-22.

———. 1966. "An Open Letter to Professor Spaeth and his Jurimetrical Colleagues." *The Journal of Politics* 28 (2):429-32.

Miceli, Thomas J., and Metin M. Coşgel. 1994. "Reputation and judicial decision-making." *Journal of Economic Behavior & Organization* 23 (1):31-51.

Mimno, David M., Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. "Optimizing Semantic Coherence in Topic Models." *Proceedings of the Conference on Empirical Methods in Natural Language Processing*:262-72.

Mitchell v. Helms. 2000. 530 U.S. 793.

New York v. Cathedral Academy. 1977. 434 U.S. 125.

Niblett, Anthony, and Albert H. Yoon. 2015. "Friendly Precedent." *William & Mary Law Review* 57:1789-824.

Oldfather, Chad M., Joseph P. Bockhorst, and Brian P. Dimmer. 2012. "Triangulating Legal Responsiveness: Automated Content Analysis, Judicial Opinions, and the Methodology of Legal Scholarship." *Florida Law Review* 64 (5):1189-242.

Owens, Ryan J., and Justin P. Wedeking. 2011. "Justices and Legal Clarity: Analyzing the Complexity of U.S. Supreme Court Opinions." *Law & Society Review* 45 (4):1027-61.

Pang, Xun, Barry Friedman, Andrew D. Martin, and Kevin M. Quinn. 2012. "Endogenous Jurisprudential Regimes." *Political Analysis* 20 (4):417-36.

Paulsen, Michael Stokes. 1993. "Lemon is Dead." *Case Western Reserve Law Review* 43:795-864.

Peccei, Christian. 2017. *Python Textmining Package* 2010 [cited January 2 2017]. Available from http://www.christianpeccei.com/textmining/index.html.

Police Department of the City of Chicago v. Mosley. 1972. 408 U.S. 92.

Richards, Mark J. 2013. *The Politics of Freedom of Expression: The Decisions of the Supreme Court of the United States*. London: Palgrave MacMillian UK.

Richards, Mark J., and Herbert M. Kritzer. 2002. "Jurisprudential Regimes in Supreme Court Decision Making." *The American Political Science Review* 96 (2):305-20.

Richards, Mark J., Joseph L. Smith, and Herbert M. Kritzer. 2006. "Does Chevron Matter?" *Law & Policy* 28 (4):444-69.

Robel, Lauren K. 1989. "The Myth of the Disposable Opinion: Unpublished Opinions and Government Litigants in the United States Courts of Appeals." *Michigan Law Review* 87 (5):940-62.

Roemer v. Board of Public Works of Maryland. 1976. 426 U.S. 736.

Rosenberger v. Rectors and Visitors of the University of Virginia. 1995. 515 U.S. 819.

Sable Communications of California v. FCC. 1989. 429 U.S. 115.

School District of Abington Township, Pennsylvania v. Schempp. 1963. 374 U.S. 203.

School District of the City of Grand Rapids v. Ball. 1985. 473 U.S. 373.

Scott, Kevin M. 2006. "Understanding Judicial Hierarchy: Reversals and the Behavior of Intermediate Appellate Judges." *Law & Society Review* 40 (1):163-91.

Segal, Jeffrey A., and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *The American Political Science Review* 83 (2):557-65.

Segal, Jeffrey A., and Harold J. Spaeth. 1996. "Norms, Dragons, and Stare Decisis: A Response." *American Journal of Political Science* 40 (4):1064-82.

———. 2002. *The Supreme Court and the Attitudinal Model Revisited*. New York, NY: Cambridge University Press.

———. 2003. "Reply to the Critics of The Supreme Court and the Attitudinal Model Revisited." *Law & Courts* 13 (3):31-8.

Sisk, Gregory C., and Michael Heise. 2012. "Ideology "All The Way Down?" An Empirical Study of Establishment Clause Decisions in the Federal Courts." *Michigan Law Review* 110 (7):1201-63.

Sisk, Gregory C., Michael Heise, and Andrew P. Morriss. 2004. "Searching for the Soul of Judicial Decisionmaking: An Empirical Study of Religious Freedom Decisions." *Ohio State Law Journal* 65:491-614.

Sloan v. Lemon. 1973. 413 U.S. 825.

Smith, Joseph L. 2014. "Law, Fact, and the Threat of Reversal From Above." *American Politics Research* 42 (2):226-56.

Smith, Joseph L., and Emerson H. Tiller. 2002. "The Strategy of Judging: Evidence from Administrative Law." *The Journal of Legal Studies* 31 (1):61-82.

Smith, Joseph L., and James A. Todd. 2015. "Rules, Standards, and Lower Court Decisions." *Journal of Law and Courts* 3 (2):257-75.

Songer, Donald R., Jeffrey A. Segal, and Charles M. Cameron. 1994. "The Hierarchy of Justice: Testing a Principal-Agent Model of Supreme Court-Circuit Court Interactions." *American Journal of Political Science* 38 (3):673-96.

Spaeth, Harold J. 1965. "Jurimetrics and Professor Mendelson: A Troubled Relationship." *The Journal of Politics* 27 (4):875-80.

Spaeth, Harold J., Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, Theodore J. Ruger, and Sara C. Benesh. 2017. "2017 Supreme Court Database."

Spriggs, James F., and Thomas G. Hansford. 2000. "Measuring Legal Change: The Reliability and Validity of Shepard's Citations." *Political Research Quarterly* 53 (2):327-41.

Stone, Geoffrey R. 1983. "Content Regulation and the First Amendment." *William & Mary Law Review* 25 (2):189-252.

Stone v. Graham. 1980. 449 U.S. 39.

Sunstein, Cass R., David Schkade, and Lisa Michelle Ellman. 2004. "Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation." *Virginia Law Review* 90 (1):301-54.

Tamanaha, Brian Z. 2009. "The Distorting Slant in Quantitative Studies of Judging." *Boston College Law Review* 50:685-758.

Tiller, Emerson H., and Frank B. Cross. 2006. "What is Legal Doctrine?" *Northwestern University Law Review* 100 (1):517-34.

Tilton v. Richardson. 1971. 403 U.S. 672.

United States v. O'Brien. 1968. 391 U.S. 367.

United States v. Playboy Entertainment Group, Inc. 2000. 529 U.S. 803.

United States v. Scott. 1999. 1999 U.S. App. LEXIS 39132 (5th Cir.).

United States v. Williams. 2006. 444 F.3d 1286 (11th Cir.).

United States v. Williams. 2008. 553 U.S. 285.

Van Orden v. Perry. 2005. 545 U.S. 677.

Wahlbeck, Paul J. 1997. "The Life of the Law: Judicial Politics and Legal Change." *The Journal of Politics* 59 (3):778-802.

———. 1998. "The Development of a Legal Rule: The Federal Common Law of Public Nuisance." *Law & Society Review* 32 (3):613-38.

Wallace v. Jaffree. 1985. 472 U.S. 38.

Wallach, Hanna M., Murray Iain, Ruslan Salakhutdinov, and David M. Mimno. 2009. "Evaluation Methods for Topic Models." *Proceedings of the 26th International Conference on Machine Learning*:1105-12.

Wallach, Hanna M., David M. Mimno, and Andrew McCallum. 2009. "Rethinking LDA: Why Priors Matter." *Advances in Neural Information Processing Systems*:1973-81.

Walz v. Tax Commission of the City of New York. 1970. 397 U.S. 664.

Ward v. Rock Against Racism. 1989. 491 U.S. 781.

Wasby, Stephen L. 2001. "Unpublished Decisions in the Federal Courts of Appeals: Making the Decision to Publish Development and Practice Notes." *Journal of Appellate Practice and Process* 3:325-42.

Westerland, Chad, Jeffrey A. Segal, Lee Epstein, Charles M. Cameron, and Scott Comparato. 2010. "Strategic Defiance and Compliance in the U.S. Courts of Appeals." *American Journal of Political Science* 54 (4):891-905.

Widmar v. Vincent. 1981. 454 U.S. 263.

Wolman v. Walters. 1977. 433 U.S. 229.

Zorn, Christopher, and Jennifer Barnes Bowie. 2010. "Ideological Influences on Decision Making in the Federal Judicial Hierarchy: An Empirical Assessment." *The Journal of Politics* 72 (4):1212-21.

APPENDIX

Document Preprocessing Steps

1) Remove common English words known as stopwords (e.g., articles, conjunctions, prepositions) as defined by the Natural Language Toolkit Python package

2) Remove a user-defined set of terms and abbreviations commonly found in judicial opinions. These terms are: 'us', 'v', 'court', '§', 'no', 'court', 's', 'd', 'case', 'end', 'footnotes', 'footnotes-', 'l', '*', 'also', 'party', 'petitioner', 'respondent', 'petitioners', 'respondents', 'appellee', 'appellant', 'appellees', 'appellants', 'id', 'supra', 'infra', 'ibid', 'whether', 'ct', 'n', 'upon', 'against', 'before', 'c', 'justice', 'U.S.C', 'cf', 'cf.'

3) Remove punctuation [$%.&#?!,":;]* and cardinal numbers [1-9][0-9]*

4) Use NLTK to lemmatize the remaining words in each opinion, which involves reducing each word to its dictionary root, to ensure that multiple variants of words are all counted as one root

LDA Model Parameters

LDA modeling is conducted using the LDA Python package. Iterations is set to 500 and random state set to 1. This package uses collapsed Gibbs sampling (Griffiths and Steyvers 2004) with symmetric Dirichlet priors. There are two Dirichlet priors: *alpha* refers to the document-topic density, and *eta* refers to the topic-word density. When the priors are symmetric, changing *alpha* results in changing the number of topics that make up each document, and changing *eta* changes the number of words that make up each topic. For this project, the symmetric priors are left at their defaults (*alpha* = 0.1, *eta* = 0.01).

Manually Identified Establishment Clause Topic Clusters

Support of schools (Parochial school funding/textbooks/Title I assistance/released time programs/curriculum issues/educational expenses tax deduction) – 1
Sabbath laws – 2
Government religious activities (School prayer/School Bible reading/graduation in a church/legislative prayer) – 3
Business regulations (Liquor licenses/ministerial exception/Fair Labor Standards Act/Unemployment benefits) – 4
Free exercise rights (Conscientious objector/compulsory school attendance/religious test/prisoner rights/right not work on Sabbath) – 5
Tax regulations (Charitable giving regulations/Sales tax exemption/Property tax exemption) – 6

Hyde Amendment (funding of abortions) – 7
Monuments (Ten Commandments/Christmas Displays) – 8
Facility access provisions – 9
Non-school funding (pro-life organization) – 10
School district lines – 11

LDA Topic Models

Content Neutrality – Thirty Topics

*Topic 0
- would say one make give time could take use think question even word statement know fact might find like two point year come first view show without way many another people thing good well want n tell put seem mean order opinion however issue call much since day believe consider
*Topic 1
- state issue question evidence judgment decision claim would constitutional opinion present see fact must find hold review federal rule could record consider decide order states make support finding ground require base law two grant remand conclude appeal appeals reverse motion first basis argument reasonable however determine judge affirm standard show
*Topic 2
- police street public person peace injunction arrest conduct officer right order clinic conviction speech demonstration protest foot leave zone breach group sidewalk cross picket violence time charge abortion evidence record protect move crowd testify hold people one enter trial building burning place white violate free t act threat entrance fact
*Topic 3
- public damage libel statement official plaintiff jury false action law malice actual liability defendant defamation defamatory private fact truth figure concern suit standard award publication conduct reputation rule article matter injury reckless evidence immunity prove proof protection falsehood hold privilege report may charge falsity individual trial punitive state verdict cause
*Topic 4
- film obscene search material warrant picture obscenity book seizure prior restraint criminal motion mail publication seize cause may matter censorship crime person magazine officer issue probable use see censor judicial exhibit police expression distribution exhibition procedure copy exhibitor determination states thing require offense item constitutionally civil violation proceeding arrest within
*Topic 5
- states power legislative purpose activity legislation subject bill would within history one act individual congressional control th effect regulation function conduct amendment make since authority st direct course legislature exercise concern could problem enact year practice respect specific influence policy seek either involve authorize time area scope therefore measure national
*Topic 6
- public right interest government may view speech political protect state free individual system freedom private matter governmental expression process citizen constitutional issue concern

official protection hold see decision claim member exercise recognize require policy subject constitutionally speak one control involve need office position action respect importance personal idea function express
*Topic 7
- broadcast cable station access television operator broadcaster channel public carry program radio system must programming broadcasting local medium licensee see regulation provision interest editorial require rule service issue license use control air time impose programmer obligation communication view common would signal carriage market newspaper provide policy number request permit carrier
*Topic 8
- candidate election political vote voter primary office ballot state general law campaign right interest require independent convention process requirement one may see voting democratic states burden support must member major petition participate nomination delegate electoral provision rule name signature candidacy place two shall party committee choose hold impose access choice
*Topic 9
- would see interest opinion even may dissent judgment ante concur however less reason decision part concern must justify could one fact apply agree simply suggest effect rather might constitutional view need indeed well protect result id majority today recognize mean believe thus join different conclude standard find least support question
*Topic 10
- employee union employer labor employment bargaining work discharge practice job service activity agency contract see patronage strike worker dismissal representative make collective majority agreement dispute management political hold policy member federal reason may assistant opinion concern support teacher unfair card use matter conduct basis prohibit condition exclusive organization benefit bargain
*Topic 11
- commercial advertising speech regulation interest advertisement information ban price restriction drug use product see may prohibit sale consumer protection market regulate advertise a prohibition service noncommercial cigarette casino promote content tobacco states health provision alcohol lottery economic order governmental contraceptive sell advance message make pharmacist activity concern physician directly require
*Topic 12
- lawyer attorney legal solicitation service professional client practice may rule fee advice profession see member investment law counsel claim person organization state litigation charitable bar business name solicit fraud conduct engage information charity public fundraiser statement potential prohibit certification speech mislead represent form advertisement interest regulation adviser cost firm license
*Topic 13
- material child obscene obscenity standard sexual minor adult community book value pornography sex offensive states interest protect person may sexually appeal indecent social prurient see find depiction video age picture game define test parent definition year sale distribution whole image young lewd patently work artistic depict statute literary explicit nudity
*Topic 14
- program fund union fee activity pay benefit cost require use expense bargaining purpose due collective charge nonmember federal support object funding agency organization provide receive

see condition compel financial assistance service political payment member assessment money expenditure litigation local shop unit issue tax may private agreement amount would year sector
*Topic 15
- tax requirement petition disclosure require see registration person name burden regulation exemption use pay purpose bank register report referendum record initiative sale information transaction financial must food impose individual circulation stamp address circulators currency provision illustration harassment reporting interest states make household exempt check challenge one account taxation classification process
*Topic 16
- statute conduct law state person prohibit criminal word apply states flag conviction construction language provision may would application construe purpose overbreadth use vague act section prosecution challenge one unconstitutional see constitutionally crime intent violate constitutional interpretation federal face prohibition vagueness statutory overbroad offense hold protect reach convict clause speech define
*Topic 17
- law right opinion freedom may power state constitutional speech free people states dissent see liberty press one must decision idea society today think judicial judge view make say protect government principle danger present protection guarantee hold judgment federal without concur process due even reason test rights men man punish every
*Topic 18
- speech v law see s eg government majority id in scrutiny burden mark the provision quotation internal omit statute ante apply challenge ban support claim judicial speaker precedent address a explain require of argument one say allow evidence point and help even provide example strict like as that issue message
*Topic 19
- press information publication newspaper public publish access news order material interest privacy medium communication report would name restraint disclosure right tape paper victim source freedom obtain article publisher reporter prior publicity id may general record issue include release confidential conversation law person need document free privilege seek editorial a see
*Topic 20
- trial jury question witness states answer criminal hearing proceeding defendant testimony privilege right prosecution investigation judge inquiry committee make grand evidence record charge indictment testify public refuse may conduct subject hold see accuse matter contempt juror conviction counsel information fair subpoena ask appear rule statement crime process prejudice give member
*Topic 21
- work copyright foreign use passport country states see alien author term travel protection citizen grant statute american citizenship year national nation right patent time trademark public fair law abroad international exist word purpose extension domain benefit authority extend provide first entry new exclusive enter limited deportation section restriction policy expression
***Topic 22 – Content Neutrality**
- speech content public forum message regulation government restriction activity expression place may property expressive access purpose speaker base use communication ban see interest viewpoint neutral based subject issue prohibition open idea manner prohibit time particular

display within general post must serve sidewalk political seek view convey provide certain category narrowly
*Topic 23
- ordinance city permit sign license adult use effect business regulation place theater sound may public door street speech one zone require secondary area regulate nude establishment licensing challenge within billboard interest time expression official discretion must community application municipality problem entertainment religious activity protect apply local cause invalid municipal control
*Topic 24
- school student religious group teacher library use speech activity religion university board public forum would college education book club institution purpose policy discrimination educational official high meeting hold access may viewpoint open child campus faculty subject academic teach see violate deny facility view parent decision organization must program particular idea
*Topic 25
- prison inmate military regulation right prisoner mail letter security official correspondence policy civilian institution base interview restriction constitutional problem permit good standard member must access within legitimate discipline write disciplinary condition rule order may allow jail authority prohibit communication officer commander visit provide open petition facility system alternative ban send
*Topic 26
- shall action provide provision right may order section person make require file hearing hold law statute agency states authority time include seek officer appeal without review relief procedure grant claim notice request judicial take constitutional violate within respect federal part exercise injunction enforcement day authorize violation term regulation reason determine
*Topic 27
- organization member membership states government overthrow activity oath association person communist force political advocate group action violence advocacy require applicant question subversive belief may evidence act find foreign mean say teach answer take admission statute communists position must character support engage conspiracy state purpose know movement loyalty security record unlawful
*Topic 28
- candidate contribution political expenditure corporation limit campaign fund money committee election corruption corporate make individual independent limitation amount support influence organization use restriction spend interest spending contributor contribute group finance disclosure may provision donor federal large receive see purpose regulation hold person direct communication b association advocacy speech appearance pro
*Topic 29
- right picketing business property state picket law private boycott public store owner secondary use restaurant center equal accommodation hold shopping plate customer commerce policy white action purpose place person product serve dispute would service company bill discrimination primary race trespass premise peaceful involve consumer labor corporation facility contract refuse member

Establishment Clause (Supreme Court) – Fifteen Topics

*Topic 0 – ***Lemon* Test**
- religious religion secular use institution purpose grant activity effect church must may organization however would advance part entanglement involve require find primary include course one political certain program building promote attempt college three make note belief although determine character excessive question year serve view group form well relationship faith line
*Topic 1
- see would fact judgment issue violate opinion make find apply reason decision hold provide even state result concur could give seek first test federal decide matter record governmental take claim agree constitutional need dissent present simply term basis rule year conclusion rather argument require reject create point quote indeed recognize
*Topic 2
- statute state states law question require hold constitutional free legislative purpose exercise provision must the might interest however opinion decision thus requirement without power benefit establish enact practice concern history prohibit problem choose impose legislature compel ground conduct legislation pass two within violation conclude determine establishment consideration mean language right
*Topic 3
- cross organization injunction transfer land stand rule private would subd relief plaintiff injury per appeal fifty cent judgment order change standing public purpose statute grant plurality observer jurisdiction action must member memorial challenge reasonable interest consider question show allege original ante property litigation requirement appeals complaint charitable register motion permit
*Topic 4
- day sale work statute shall law may rest business license person exception section week labor provide sell one activity legislation permit regulation open time find ban retail hour make class employee people keep necessity effect except prohibit act closing observance provision operation six legislature within set employment c public commodity
*Topic 5
- one see may would government law principle even say dissent respect think today opinion view majority well without could word believe general much another good seem come way true make among example write like show course long involve ante place though call present two early become less matter might know
*Topic 6
- aid school religious program government private student see a religion choice provide fund benefit sectarian chapter direct neutral receive voucher public recipient individual majority plurality secular support indoctrination educational pervasively funding education concur neutrality equipment ante criterion tuition diversion material money participate employee choose number evidence interpreter available post content
*Topic 7
- purpose id religion creation text science secular evolution evidence commandments theory religious bill teach see legislative school belief document teaching history scientific advance legislature statement freedom legislator even life ten action objective particular curriculum law

original include support historical academic counties nothing student resolution doctrine classroom two concur omit view
*Topic 8
- school public child provide state sectarian program teacher education service parent student nonpublic instruction parochial private textbook educational attend aid teach pupil secular use fund law loan material test teaching class book district system effect shall reimbursement subject church elementary function provision statute secondary authorize assistance cost may local expense
*Topic 9
- student school speech forum group public policy religious activity use club would viewpoint open meeting facility access permit content private graduation message speaker university endorsement may deny limited hold discrimination allow subject high meet purpose district majority class worship could endorse child must particular view exclude organization property expression election
*Topic 10
- tax payment exemption taxpayer benefit church property government fund pay subsidy deduction amount tuition money income use contribution institution expenditure cost support state facility credit receive exempt provide sale activity service charitable revenue audit claim burden contribute form grant bill federal financial part direct generally taxation aid establish provision thus
*Topic 11
- religious exemption belief war religion accommodation conscientious exempt benefit particular conscience objector district objection claim burden service group individual moral would states person participation military exercise village whose organization training faith community basis certain government rather conviction governmental special personal duty right nonprofit reason unemployment oppose political nonreligious neutral form
*Topic 12
- prayer display government religious religion practice message creche endorsement public holiday symbol concur part city see official belief menorah test town particular convey ante observer endorse chaplain faith history invocation secular christian post tree place monument context many community silence view member tradition a effect historical sign include celebration reasonable
*Topic 13
- religious religion public establishment state church exercise right states support people freedom shall every make time practice sect power free liberty establish authority give school may must history worship say individual group faith force belief part take many education society form american child man country separation first prohibit others rights
*Topic 14
- abortion fund woman necessary funding service life interest health medically pregnancy grantee right medical federal adolescent may protect constitutional benefit protection choice governmental freedom program participate childbirth recognize pregnant mother cost potential eligible equal carry provide care legitimate pay choose indigent include constitutionally involve liberty term conclude class counseling restriction

Employment Discrimination – Fifteen Topics

*Topic 0
- would see make one even decision could individual apply opinion reason question law view must ante say dissent part basis however give clear concern the majority mean rather different effect might statement respect well issue point employee first base suggest discrimination id consider two conclude argue example application simply rely
*Topic 1
- job fact find standard may test require position use finding discriminatory issue must without judgment requirement thus record work establish determine appeals deny question order person make reject evidence take number remand member opportunity show however necessary general two since present employ one hold result applicant policy performance fail reverse
*Topic 2
- provision discrimination employment statute employee language employer provide history section would legislative amendment bill intend practice interpretation term shall religious enact intent prohibit e person require amend statutory purpose find include construction legislation work congressional meaning add sex title contract leave id remark take interpret day religion contain reasonable prohibition
*Topic 3
- seniority system employee job discrimination employment practice discriminatory employer line rule decree effect hire company right act injunction unlawful layoff would union white driver hiring black prior hold city work minority victim past bona race preliminary provide offer consent status date time fide states application present discriminate transfer apply claimant
*Topic 4
- employer employee claim s employment v act harassment discharge action supervisor discriminatory conduct hold liability retaliation discrimination environment constructive work hostile sexual see plaintiff rule take mark eg liable internal quotation allege defense condition agency actionable decision reasonable standard term policy may base workplace scope authority adverse jury judgment matter
*Topic 5
- charge file day period limitation practice time within agency employment notice filing allege e unlawful act date requirement occur proceeding timely action discrimination may complaint see employer information must local year state suit purpose investigation tenure provision enforcement require intent allegation deferral present shall statute complainant terminate violation termination commence
*Topic 6
- school racial race student white plan district black methadone teacher use classification board interest dissent s segregation v program in percent assignment public year local based state opinion post a child high user conscious achieve minority educational history plurality eg seek american system education desegregation racially basis scrutiny see population
*Topic 7
- age retirement worker old plan employee benefit year employer pension mandatory service young pilot provision individual fide bona provide basis retire purpose business defense flight subterfuge exemption engineer qualification system officer see th statute policy position disability captain involuntary firefighter become cost state evade status permit arbitrary normal eligible term

*Topic 8
- pay sex benefit employee plan equal contribution employer woman wage pension base men insurance male female compensation based payment work cost decision differential rate retirement examiner fund shift annuity company defense salary provide hold make require night year basis retroactive amount life offer receive violation individual table condition option violate
*Topic 9
- relief class award backpay employee remedy claim order contract damage agreement would bargaining policy collective retroactive may arbitration union right member action individual equitable plaintiff appropriate pay discretion statutory make hold a reinstatement date liability appeals available provision power see however interest remedial question judgment grievance conduct circumstance deny offer
*Topic 10
- alien states citizen law citizenship officer country position class classification interest state may hold foreign resident right practice public within requirement police national person status community apply lawfully civil become political statute permanent scrutiny subject service control admit employ immigration bar lawyer function corporation broad japanese american perform residence work
*Topic 11
- pregnancy woman employee disability pregnant sex leave work policy benefit employer female treatment male related job discrimination condition basis treat risk require plan worker program coverage clause medical v ability similar state employment law child provide gender exclusion men physical childbirth protection would week guard second inability exclude business neutral
*Topic 12
- state power regulation employment rule public government states constitutional federal see may authority program equal judge protection exercise interest require statute service official person provide violate policy duty exception law act fund within exclusion therefore limit local level provision private political legislation responsibility particular three legislative serve governmental certain officer
***Topic 13 – *McDonnell Douglas* Test**
- employer plaintiff evidence discrimination disparate employment burden practice impact decision reason prove prima facie defendant show proof treatment action see claim factor applicant legitimate may must hire establish require individual cause e sex plurality race nondiscriminatory rule shift hold statistical jury intentional pretext motive neutral motivate liability judgment business direct
*Topic 14
- federal right state action law claim agency decision statute private states administrative suit discrimination cause remedy may civil proceeding see hold jurisdiction judicial complaint rule bring give provide issue review judgment trial effect opportunity enforcement sector district sue procedure determination grant present dismiss section create applicable merit waiver violation plaintiff

Establishment Clause (Circuit Courts) – Twenty Topics

*Topic 0
- would see majority one opinion even make religious may fact decision government could constitutional require question public issue view dissent n concur rather way must reason take hold two say emphasis part apply time however indeed suggest analysis like give consider record simply without well point clear believe might context
*Topic 1
- states national military section power country chaplain motto entry alien officer statement authority policy interest constitutional visa promotion would board provide attendance security order review action government foreign judicial basis executive provision base discrimination likely nation deny purpose chapel find ban requirement challenge congressional future exercise individual statute facially reason
*Topic 2
- policy prayer graduation student ceremony message invocation state event game endorsement song exercise speech deliver benediction official vote endorse high coercion speaker football public participate religion coerce senior content participation music permit choose practice allow team theme hold would class select unconstitutional graduate decision volunteer control nonsectarian purpose performance member
*Topic 3
- school student public district child teacher activity parent class classroom time education find program teach board require high religious attend instruction curriculum course participate would elementary attendance hold educational policy secondary state take bible meeting material free may use authority related hour day must system subject grade however permit principal
*Topic 4
- speech forum religious group public policy access activity use would viewpoint permit government open allow organization private discrimination facility content club meeting may limited event interest equal speaker restriction exclude purpose create expression free religion deny violate subject neutral community school right student hold worship exclusion grant reasonable must property
*Topic 5
- statute prayer school purpose silence moment public amendment exercise word state legislative student law recitation shall pledge bill day teacher unconstitutional may silent secular flag recite period activity child history pray add one legislature right version evolution amend nation minute belief enact phrase provide hold daily engage legislator section theory
*Topic 6
- cross stand plaintiff injury seal standing use symbol see memorial challenge government religious th city hold federal allege resolution public action war religion effect contact taxpayer suffer message park violate fact establishment direct state land confer must christian display one sufficient district county belief member community claim lack plaintiffs conduct
*Topic 7
- coverage contraceptive religious plan provide health accommodation burden organization government insurance plaintiff self employer mandate require insured exercise see opt substantial regulation interest would profit belief third law employee service administrator non scheme objection eligible woman must requirement use form group cost pay exemption beneficiary substantially corporation party impose contraception

*Topic 8
- religious religion church belief establishment state clause worship rule may use member group practice note service faith christian question free course include institution view one conduct principle doctrine well form nature life basis deny matter law make prohibit action activity concern grant teaching support judgment general respect seek decision constitute
*Topic 9 – Business Regulations
- use property land public church would government area city sale building facility permit lease ordinance private space park water provide withdrawal within district special plan regulation include organization zone site agency construction legislative project rent purpose application year locate foot quality standard action development operate provision shelter institution new housing
*Topic 10 – Legal Terms
- right exercise law burden free interest religion religious state government federal claim hold see violate apply protection statute constitutional regulation may protect require compelling belief th impose prison substantial official conduct power neutral general prohibit act inmate authority practice person within scrutiny immunity establish prisoner must substantially condition requirement compel
*Topic 11
- prayer practice legislative meeting invocation faith public one sectarian government content session offer chaplain belief open member board reference christian town body tradition give deliver religious state opportunity pray particular official advance policy name coercion district legislator participate hold citizen stand legislature historical ask proselytize nonsectarian speaker record commissioner history
*Topic 12
- district claim th plaintiff judgment action issue state see appeal grant motion violate review summary deny order defendant fact find violation determine file decision allege complaint must evidence conclude hold consider fail request conduct two argument make however policy relief standard argue dismiss apply challenge question regard reason without although
*Topic 13
- display holiday creche public symbol government menorah message endorsement religious city sign secular observer tree reasonable erect scene private religion endorse season park place painting state convey nativity context district find celebrate disclaimer permit foot fact use light building large concur test celebration area unconstitutional part year unattended location forum
*Topic 14
- injunction plaintiff would right preliminary say t show relief one defendant law officer make merit could may cause harm judge act n attend order ask likelihood injury tell damage enjoin success irreparable sanction day appeal know trial abuse part violation evidence available counsel others testify sexual give get person action
*Topic 15
- display monument purpose religious secular observer government history commandments reasonable endorsement historical religion ten test message plaque context text law note document would courthouse effect american th concur evidence include view endorse objective symbol fact prong legal find see state county action part contain place conclude hold erect significance statue

*Topic 16 – ***Lemon* Test**
- religious religion purpose secular state test effect advance government primary entanglement must prong establishment practice first excessive find use violate see governmental note challenge may part law however day neither second promote thus challenged foster therefore three statute neutrality benefit provide particular good serve establish endorsement satisfy analysis require unconstitutional
*Topic 17
- employee religious church employment employer would discrimination work belief statute exercise union require minister hold pay practice institution jurisdiction accommodation member entanglement worker charge free clause apply interest discharge amendment teacher position exemption duty e exception section th application ministerial determine faculty hardship burden government client deny labor sex relationship
*Topic 18
- program school service provide sectarian aid state fund private religious institution public parochial child benefit education see statute cost funding educational secular pervasively college nonpublic taxpayer use would provision special pay grant direct religion attend government neutral available issue chapter choice bond expenditure agency parent result assistance tax require receive
*Topic 19
- tax section law exemption benefit organization statute provision individual care requirement kosher charitable exempt payment contribution make provide power food income receive religious challenge purpose deduction legislative medical find taxpayer authority thus church fund interest service effect require states statutory product ordinance federal see license sect governmental member pay solicitation