FINDING LAW: LEARNING IN THE JUDICIAL HIERARCHY

Deborah Beim

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF POLITICS

Adviser: Charles M. Cameron

September 2013
Abstract

The U.S. Supreme Court supervises the disposition of hundreds of thousands of cases each year, and the law that governs these disputes is not static: new problems appear, new classes of disputes arise, and doctrine must adapt to govern their resolution. This dissertation explores how the Supreme Court learns to create and extend doctrine to adjudicate these cases. I argue the Supreme Court relies on the Courts of Appeals as laboratories of law, observing their decisions and reviewing those that best inform legal development.

I develop a formal-theoretic model in which the Supreme Court supervises two lower courts deciding similar cases. A tension arises because the Supreme Court can see only case outcomes from the lower courts—only upon costly review can the Supreme Court determine whether the arguments that support a lower court’s decision are sound. But, because an unbiased judge only makes an extreme decision when there is an imbalance in the parties’ arguments, the Supreme Court is able to draw some inferences without review. The Court thus leverages multiple Courts of Appeals decisions to identify which will be most informative to review. Upon review, it learns from the arguments presented in the instant case to develop legal doctrine.

The model generates a number of empirical predictions, which the subsequent chapters test. First, because moderate decisions are ambiguous and therefore more informative to the Supreme Court, the model predicts the Supreme Court should be more likely to review these. The data support this prediction. Second, consistent with the model’s predictions, when the Supreme Court is resolving a conflict between lower courts, it is more likely to review cases on the side whose doctrine it ultimately rejects, especially when lower courts are heterogeneous. Finally, the model predicts the Supreme Court will reverse extreme decisions more than moderate decisions; empirically, the Court is more likely to reverse liberal decisions than moderate ones, but the same is not true for conservative decisions. The results shed light on
how hierarchy eases the inherent difficulty and uncertainty of crafting law and on how the
Supreme Court learns to create doctrine.
Acknowledgements

This is a dissertation about how to learn by aggregating the knowledge of others—how to identify sound advice, how to reconcile conflicting suggestions, and how to make a decision that is more than the sum of its parts. I was fortunate enough to spend graduate school surrounded by brilliant people eager to share their knowledge and experience this for myself. My committee read my papers time and again; watched the same talks repeatedly, each time asking more difficult questions; and met with me whenever I asked. They—Chuck Cameron, Brandice Canes-Wrone, John Kastellec, and Adam Meirowitz—were individually excellent. Collectively, they were unparalleled. I cannot imagine greater fortune than to have been advised by this group. In their most extraordinary moments, they talked about my project of their own initiative and without me there. At the time that was terrifying and infuriating, but absolutely nothing could have been better for my work or my professional development. For their unified dedication to my success, I offer sincere thanks.

Chuck Cameron is a uniquely great advisor. Good advisors are generous with their time, which Chuck always was, but he was also generous with his intellect, thinking deeply about my ideas then sharing his insights. Nothing could have been more humbling or helpful. Over the past five years, Chuck was aware of my progress even when I was not, kept a permanent longview of my career, and guided me to success. His enthusiasm for understanding institutions is infectious, and his perspective on political science—broader and more holistic than I could have imagined—is singular and invaluable. His advice turned my research into something I could be proud of. Chuck also shared his office space with me, which was instrumental to my success. My profound thanks to him for having advised me with such dedication, generosity, and cheer.

I learned how to do research by watching John Kastellec work. For this I am impossibly
grateful: were it not for the countless hours I spent in his office, I would not know how to do political science. Every graduate student needs a Yoda, and he was mine. John was a constant source of advice and instruction, infinitely patient with my endless questions and generous with his time. Although his standards often seemed unreachably high (and still do), trying to meet them improved my writing and clarified my thinking. His kindness and good nature made co-authoring and being his advisee a pleasure, and I’m deeply thankful to have had the opportunity to work with him.

Brandice Canes-Wrone offered advice that was clear and immediately constructive; her perspective on the discipline is invaluable and her help in framing my work made all the difference. I could enter Brandice’s office with cloudy ambitions for a distant project and leave with a precise understanding of what to do and how to do it. She made this project a reality, and I am very thankful for all the help she gave me.

Since our first meeting at recruitment, Adam Meirowitz supported my professional goals with superlative capacity and humor. Adam helped me to construct the model in this dissertation, then encouraged and guided me when I feared mistakes. The reassurance and education I received from him were critical to my success and the success of this project; most meaningfully, Adam taught me to build models I could solve instead of helping me solve models. His friendly curiosity made work fun, and it was an honor to study with him.

Tom Clark, Cliff Carrubba, David Klein, Bill Landes, Stefanie Lindquist, Richard Posner and John Summers were very generous to share their data with me, which made this project possible. I am very grateful to them. Parts of this dissertation were presented at seminars at the University of Chicago Harris School of Public Policy, Dartmouth College, Emory University, George Washington University, Georgetown University, the University of Illinois at Urbana-Champaign, the University of Michigan, the Stanford University Graduate School of Business, SUNY Stony Brook, Vanderbilt University, Yale University, the 2012 Midwest
Political Science Association conference in Chicago, Illinois, and the 2012 American Political Science Annual Meeting in New Orleans, LA. I am deeply thankful for the comments and questions I received at each of these presentations.

I also wish to thank Lewis Kornhauser, who shaped the way I think about courts. His wisdom on modeling and understanding courts in their most abstract and pure sense has been invaluable, and it has been a great honor to work with him. Keith Whittington, who served as an examiner of this dissertation, provided immensely constructive and insightful feedback, and I’m very grateful for the time he spent reading and thinking about this work. Many thanks to Tom Clark, who was an excellent role model and read my papers for no reason but his own generosity. Thanks also to John Londregan, for introducing me to “Advocates;” to Matias Iaryczower and Nolan McCarty, whose enthusiastic seminar participation taught me how to give a talk and how to enjoy it; to Alex Hirsch, who served as a terrific mentor; to Michele Epstein and CSDP for their support; to my great colleagues, who made Princeton and grad school tolerable in their worst moments; to my dear friends who proofread, consoled, and cheered me on from afar; and to Kosuke Imai, Kris Ramsay, and the rest of my hallmates. Every workplace should be as inspiring and joyful as mine was. In How I Work, Hal Varian suggests students talk to their pets about their models. I don’t have a dog, so I talked to Mack and Montana. My thanks to them for listening, and to them and their owners for easing my stress in its worst moments.

My parents, Mark Freeman and Gina Beim, boundlessly and eagerly offered their love and support throughout my life, my education, and my time in graduate school. They made it possible for me to succeed and shaped my understanding of success. I am deeply thankful for all they’ve done and for the grace of having such outstanding parents. Finally, thanks to my grandmother, Ellen Kopelman. She wanted very much to see the end of this project, and I hope the result would have made her proud.
## Contents

1 **Introduction: Finding Law**  
   1.1 The Judicial Hierarchy ................................................. 4  
   1.2 Structure of the Dissertation ....................................... 10  

2 **Theory**  
   2.1 The Model ........................................................................ 14  
      2.1.1 Decision making in the lower courts ................................ 15  
      2.1.2 Learning and decision making at the Supreme Court ............ 17  
      2.1.3 Preferences and beliefs .................................................. 19  
   2.2 Optimally Learning from Agents’ Decisions .......................... 21  
      2.2.1 Supervising two unbiased judges .................................... 22  
      2.2.2 Bias in the lower courts ............................................... 26  
      2.2.3 Case consolidation ....................................................... 30  
   2.3 Discussion and Conclusion ............................................... 31  

3 **Review**  
   3.1 Theory and Hypotheses .................................................. 44  
   3.2 Data .................................................................................. 45
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3  Panel Composition and Decisions</td>
<td>48</td>
</tr>
<tr>
<td>3.4  Results</td>
<td>49</td>
</tr>
<tr>
<td>3.5  Discussion and Conclusion</td>
<td>52</td>
</tr>
<tr>
<td>4    Conflict</td>
<td>56</td>
</tr>
<tr>
<td>4.1  Measuring Doctrine</td>
<td>58</td>
</tr>
<tr>
<td>4.2  Lower Court Conflict and Supreme Court Review</td>
<td>60</td>
</tr>
<tr>
<td>4.3  Theory and Hypotheses</td>
<td>61</td>
</tr>
<tr>
<td>4.4  Data</td>
<td>64</td>
</tr>
<tr>
<td>4.5  Analysis</td>
<td>66</td>
</tr>
<tr>
<td>4.6  Results</td>
<td>68</td>
</tr>
<tr>
<td>4.6.1 Strategic litigants</td>
<td>73</td>
</tr>
<tr>
<td>4.7  Discussion and Conclusion</td>
<td>75</td>
</tr>
<tr>
<td>5    Reversal</td>
<td>78</td>
</tr>
<tr>
<td>5.1  Understanding Reversals</td>
<td>79</td>
</tr>
<tr>
<td>5.2  Theory and Hypotheses</td>
<td>82</td>
</tr>
<tr>
<td>5.3  Data</td>
<td>83</td>
</tr>
<tr>
<td>5.4  Analysis</td>
<td>86</td>
</tr>
<tr>
<td>5.5  Discussion and Conclusion</td>
<td>92</td>
</tr>
<tr>
<td>6    Conclusion</td>
<td>95</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction: Finding Law

In November 2012, the Supreme Court heard the case of Vance v. Ball State University. Maetta Vance claimed to have been subject to sexual harassment at work and was bringing a Title VII suit against her employer, Ball State University. Her claim turned on the legal question of whether the harasser was her “supervisor” or merely a coworker. The Seventh Circuit, where Vance’s claim was brought, defined a supervisor narrowly—including only those with the power to “hire, fire, demote, promote, transfer, or discipline” (See 646 F.3d 461 and 276 F.3d 345). The First and Eighth Circuits had similar rules. In contrast, the Second, Fourth, and Ninth circuits (as well as the EEOC and the United States Attorney General) defined a supervisor as one with “authority to direct and oversee their victim’s daily work” (Goldberg and Ortiz 2012).

It fell to the Supreme Court to resolve the conflict between the circuits, and choose one definition—but they had to do so with relatively little information, for the justices had little to no experience being supervisors or supervised in a workplace like Vance, who worked as a caterer. Their questions during oral argument sought to understand the complexities of working in a kitchen like Vance’s, asking whether a more senior employee who controls the
music a supervisor—is subjecting the harassed to unpleasant music a form of power? Is the control of tasks (chopping onions instead of celery, for example) a form of power worthy of being called “supervisor?”

To clarify the definition of a supervisor, and resolve the conflict between circuits on the appropriate definition, the Supreme Court chose to review the decision of the Seventh Circuit. Upon hearing the case, the justices expressed frustration to the attorneys that the Seventh Circuit’s doctrinal position was not being defended. The counselors—all of them, including the amici and the counselors for both petitioner and respondent—failed to defend the Seventh Circuit’s definition of a supervisor. Upon realizing there were no arguments to support such a definition, Justice Alito suggested that perhaps the issue should return to the lower courts, so that arguments could be developed further.¹

Contemporary judicial politics would have much to say about such a case. Driven by the seminal work by Segal and Spaeth (Segal and Spaeth 2002, Epstein et al. 2007), scholars have understood Supreme Court justices to be political actors whose decisions are driven primarily by their personal ideologies. But the attitudinal model, in its original conception, argues only that justices’ decisions on the merits are guided by ideology. Whether Ball State or Maetta Vance will be vindicated is driven by the justices’ preferences, according to that theory and to an abundance of supporting empirical evidence. But the resulting doctrine—which will guide the disposal of future, similar cases—is a far more complex object, which is shaped by far more than preferences alone.

Compared to the Supreme Court’s disposition, doctrine has received somewhat less attention. Although political science has begun to address the sources and content of doctrine (see e.g. Maltzman, Spriggs and Wahlbeck 2000, Corley, Collins and Calvin 2011, Carrubba

¹The text of the Supreme Court’s oral argument in this case is available at http://www.supremecourt.gov/oral_arguments/argument_transcripts/11-556.pdf
and Clark 2012, Clark and Carrubba 2012, Clark and Lauderdale 2012), it remains unclear exactly what the Supreme Court seeks to accomplish when it hears a case. Those studies that have considered doctrine as a political object have tended to focus on discretion and precision in the law—that is, on whether to articulate a rule or a standard—or on an ideological disagreement between the Supreme Court and other branches of government that leads the Court to moderate its most-preferred position. Few, however, have focused on understanding the complexities that arise because of uncertainty. The claim I make in this dissertation is that the Supreme Court seeks to inform itself to enable the creation and development of legal doctrine. Acknowledging that justices have individual preferences, and that they are constrained by institutional requirements, I claim what justices primarily do is search for the correct combination of words to express a rule they have chosen. In other words, they seek a written opinion that will translate into their ideal resolution of current and future disputes.

At its root, the Vance case is about defining a legal term—about the categorization of some employees as co-workers and others as supervisors. Indeed, the bulk of the Supreme Court’s work comes in such tasks—defining houseboats as houses or cars (Lozman v. City of Riviera Beach), defining antiquated legal terms like defalcation (Bullock v. BankChampaign, N.A.), and defining who is and is not a supervisor. For the justices, these questions contain large amounts of uncertainty. Even those who know how they would dispose of Vance’s case do not know how they would dispose of all similar cases. What is the threshold beyond which a person is a supervisor? As Justice Roberts asked during oral argument, “Do you have any idea how this works on the ground when people complain about the exercise of authority by a coworker who has specific responsibilities that might be reviewed as supervisory?”

This dissertation explores how the Supreme Court answers these difficult questions—how it learns to make and develop legal doctrine in the face of uncertainty and incomplete
information. The answer, I argue, is reliance on lower courts. The Supreme Court relies on lower courts to filter lawyers’ arguments, identifying valid points and distilling them into legal opinions. The opportunity to learn from subordinates’ successes and failures is one of the fundamental strengths of hierarchical organizations. American states are referred to as laboratories of democracy for just this reason: “It is one of the happy incidents of the federal system that a single courageous State may, if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country” (New State Ice Co. v. Liebmann, 285 U.S. 262 (1932), Justice Brandeis, dissenting). The federal government can observe states’ social and economic experiments and adopt the best practices. The same is true in the federal courts, where new law is developed in the lower courts as the Supreme Court watches. Inferior courts filter arguments for the Supreme Court, identifying doctrines that are best for particular areas of law. This hierarchy of experimentation can help the judges at the top develop informed opinions and make good decisions. In short, hierarchy can help superiors learn. But that learning is not always straightforward. Aggregating the results of many agents’ experiments, and understanding the causes of their successes and failures, requires careful supervision and strategic review. In this dissertation, I explore how a supervisor can best learn from a group of agents in the context of the federal judicial hierarchy. I show how the Supreme Court uses the Courts of Appeals as laboratories of law, observing their decisions and reviewing cases to learn about doctrine.

1.1 The Judicial Hierarchy

The Supreme Court sits at the top of the American judicial hierarchy, supervising two inferior tiers of federal courts—the District Courts and the Courts of Appeals—as well as state courts. As in many hierarchies the federal judiciary has developed a division of labor, wherein judges in different tiers have functionally different roles. This specialization has
a profound effect on how, and how well the judiciary functions (Cameron and Kornhauser 2005). Therefore, it is useful to begin by briefly describing the structure of the federal judicial hierarchy.

Most federal cases begin in one of 94 federal District Courts. District Court judges sit alone and hear many cases a day, most of which are legally quite straightforward (Glick 1988, Jacob 1983, Domnarski 2009). In all, the judges dispose of about 350,000 in a typical year (Administrative Office of the United States Courts 2013), but much of a district court judge’s time is spent settling cases (one result of which is that they hear even more cases than they dispose). These judges serve as implementers of law, and although that may allow discretion, it is almost entirely discretion in doctrinal application only Baum (1997), Goldman and Jahnige (1976). Losing litigants may appeal to the Courts of Appeals; if they do, the appellate court is required to hear their case.

The Courts of Appeals consist of 179 judgeships divided geographically into 12 circuits. In each circuit, panels of three judges are composed to hear cases, so that over the course of the year a judge will sit with many different colleagues. The Circuit Courts decide about 65,000 cases in a typical year. Appellate court judge Coffin (Coffin 1980, 94-5) defines three types of cases that comprise his workload: those that require him to correct error and preserve continuity and consistency in the law, those that require traditional agency review, and those in which the judge influences the direction of the law. Appellate judges thus often write and publishing an opinion that explains the legal rationale behind their decision—and that becomes legal precedent for the district courts. In fact, much of the duty of a Courts of Appeals judge is to serve as a supervisor to the district courts, developing doctrine for the circuit and ensuring uniformity between the district court judge’s decisions. Thus, this is a tier of middle management—while the Courts of Appeals are principals to the District Courts, they are agents of the Supreme Court.
By comparison, Supreme Court justices decide very few cases: in a typical year, they issue about 80 decisions. Most commonly these are cases that lower courts have already heard, but one or more parties in the case have filed a petition for *certiorari*, asking for the Supreme Court to hear the case. About 10,000 petitions for *cert* are filed in any given year; of these, about 1,000 are placed on a “discuss list” of cases that are viable candidates for review (Schoen and Wahlbeck 2006). At this stage, the justices have little information about a case: their clerks have read a brief from the petitioner and written a brief summary memo; if the respondent filed a brief, or if outside parties filed supplemental *amicus curiae* briefs in favor or against taking the case, the justices and their clerks have those as well. All of these documents, however, are relatively short—the justices may have an expectation of what questions will arise and what arguments will be presented if the case is heard, but they will not know for sure. With this little information in hand, the justices briefly discuss these cases and vote on which to hear. The case is granted review if four of nine justices vote in favor of hearing a case. About 80 petitions for *cert* are granted each year; the rest are denied review and remain in their prior state.

The Supreme Court reviews so few cases because it “casts itself in an Olympian role” (Shapiro 2006): while lower courts focus on dispute resolution, the Supreme Court focuses on articulating doctrine—that is, on structuring dispute resolution by crafting rules that apply to sets of cases. As such, a natural question is how the Supreme Court chooses what doctrine to articulate. Because the task of articulating doctrine involves inherent uncertainty (Black and Owens 2012), understanding the creation of doctrine requires an understanding of how the Supreme Court learns. To focus on the theoretical answer to such a question, a number of papers have imagined a judiciary that consists only of one judge, who hears all cases, alone—without colleagues and without inferior or superior courts (Cooter, Kornhauser and Lane 1979, Niblett 2013, Baker and Mezzetti 2012). Like the model presented here, those
often suggest that a judge’s time is best spent on most informative cases.

The Supreme Court’s role, however, it not only to articulate doctrine but to do it from atop a hierarchy. In a hierarchy where multiple agents communicate to principals, agents’ messages can interact with—and sometimes counteract—one another (Epstein 1998, Dewatripont and Tirole 1999, Battaglini 2002, Minozzi 2011). Learning from many agents thus provides more information than the sum of each’s message. However, the judicial hierarchy affords the Supreme Court two benefits: it can aggregate the decisions of lower courts and, if it wishes, it may review some of their decisions to better understand them. Which cases deserve further review? Calvert (1985) considers a principal who has two potential sources of advice and can choose to learn from only one; however, the principal does not observe anything before choosing which advisor to consult. In the judiciary, the Supreme Court sees certain salient facts—like who made the decision, and what decision was reached—before deciding whether to review a case. This is important in understanding the particularities of judicial learning.

Because the Supreme Court takes so few cases, understanding which of these many Courts of Appeals decisions deserve Supreme Court review, and how to reconcile the inevitable differences that arise between them, is a critical task. The question has received immense attention from judicial politics scholars (e.g. Cameron, Segal and Songer (2000), Lax (2003), Kastellec (2007), Clark (2009), Beim, Hirsch and Kastellec (2012)). Most of this research has understood the hierarchy as a disciplinary organization; thus, the advantage of learning from subordinates is generally ignored to focus on the difficulty of monitoring them. This framework prioritizes uniformity in the law and seeks to understand how the Supreme Court can ensure consistency in the large number of decisions made in the lower courts—not to mention compliance with their own preferences. Furthermore, with few exceptions, these models are dyadic, that is, the Supreme Court supervises only one lower court, whereas in
reality the Supreme Court supervises multiple lower courts simultaneously (Lindquist, Haire and Songer 2007). A small set of models consider the implied dynamics of supervising many inferior courts: in these “tournament” models of the judiciary lower courts compete to be least non-compliant (Cameron 1993, McNollgast 1995, Lax 2003). Still, because they are focused on discipline rather than learning, these do not consider how lower courts’ decisions can be useful. All find that the Supreme Court should prefer to review decisions of which it is suspicious, such as decisions made by courts it sees as biased and decisions with which it believes it is likely to disagree.

But the Supreme Court often chooses to grant certiorari—that is, to review—the decision of an ideological ally. Lindquist, Haire and Songer (2007) and Walson (2011) both show that while the Supreme Court reviews decisions from ideologically opposed lower courts more often than allied lower courts, it still reviews its allies at a significant rate. The failure to explain such a phenomenon points to a weakness of existing theory on judicial supervision—rather than reviewing only the lower courts of whom it is suspicious, as most theories have proposed, the Supreme Court often reviews decisions with which it is likely to agree. Similarly, research also shows that justices prefer to review the decisions of high-quality judges. Perry (1991) writes of his interviews with clerks, “Contrary to what I had expected, justices would rather take a case where the opinion below is from a well-respected judge, because they will start from a more informed point.” Such a pattern suggests that rather than monitoring their subordinates, the Supreme Court uses lower courts’ decisions instrumentally.

Indeed, scholars of the judicial hierarchy have begun to pay more attention to the creation and development of good legal doctrine, particularly through the use of lower courts. For example, lower courts’ citation practices have been shown to inform the Supreme Court about how doctrines have been interpreted (Hansford, Spriggs and Stenger 2010) and language from lower courts’ opinions often finds its way into the opinions of the Supreme Court.
Collins and Calvin (2011). In fact, most Supreme Court opinions cite at least one Courts of Appeals opinion other than that of the case they are reviewing (George and Berger 2005). That is to say, the Supreme Court is frequently aware of, and informed by, multiple lower courts’ decisions.

Importantly, these decisions informing the Supreme Court are often in conflict with one another. A particularly robust pattern in certiorari decisions remains theoretically under-explored: the Supreme Court is most likely to grant certiorari in order to resolve a conflict between two lower courts. Rule 10 of the Rules of the Supreme Court of the United States mentions conflict in the lower courts as a reason to consider granting certiorari, and indeed, conflict is an excellent predictor of review (see Caldeira and Wright (1988) and Estreicher and Sexton (1984)). Black and Owens (2009) show that conflict is a better predictor of review than is lower court ideology. That a relationship between conflict and learning exists has long been recognized, but there has been much less work offering explanations for said relationship (though see Clark and Kastellec (2013)). Empirical patterns suggest that the Supreme Court uses conflict to its advantage—for example, when lower courts are in disagreement, the Supreme Court decides in favor of the side that more circuits agree with (Klein and Hume 2003, Lindquist and Klein 2006). When lower courts encounter new areas of the law, the Supreme Court adopts their rules after allowing them to percolate sufficiently (Klein 2002).

Still, the mechanism by which lower courts facilitate doctrinal development remain somewhat under-theorized. Carrubba et al. (2012) develop a theory of how the Supreme Court could adopt rules close to its own ideal through strategic review. Clark and Kastellec (2013) describe how the Supreme Court observes a sequence of decisions in the lower courts to learn about law. This dissertation joins these studies in understanding how the Supreme Court benefits from the efforts of its subordinates. (Outside the judiciary, this has received thorough attention—see Garicano and Zandt (2013) for a review). In particular, the focus
of this dissertation is on learning from subordinates. Building off these findings, the dissertation focuses on how the Supreme Court learns from the decisions of lower courts in order to construct new legal doctrine. Specifically, the theory describes how the Supreme Court aggregates lower court decisions to learn first which case will be best to review, and second, what decision to make upon review. In so doing, the theory speaks to scholarship on strategic communication in hierarchical organizations in general, and to long-standing literatures on the judicial hierarchy in specific.

1.2 Structure of the Dissertation

In Chapter 2, I present a formal model of learning in hierarchical organizations. This model concretizes the theory of how, and what, the Supreme Court learns from lower courts. The model makes a series of novel predictions, including which lower court case will be most informative and what the probability of reversal will be conditional on review. The following chapters test each of these predictions.

In particular, the theory suggests moderate decisions—where each party prevails on some counts—provide much more information than decisions where one party prevails on all counts. If the Supreme Court is attempting to learn, and to develop doctrine, it will be more likely to review these cases. Chapter 3 uses a dataset of about 6,000 Courts of Appeals cases to show that, as predicted, the Supreme Court is more likely to review moderate lower court decisions—decisions where each party wins on one count—than extreme lower court decisions—decisions where one party wins on all counts.

Chapter 4 examines the Supreme Court’s doctrinal choices—those legal pronouncements whose impact is most profound. That chapter shows that the Supreme Court is far more likely to reject the doctrines it chooses to review—a result which is predicted by the formal model. The results suggest that the Court is not primarily concerned with ensuring that
lower courts follow established doctrine—that is, with deciding easy cases as the Supreme Court would like—but rather with the establishment and extension of new doctrine—that is, with the development of new rules for disposing of future cases. Finally, Chapter 5 turns to the Supreme Court’s dispositional votes—which cases it chooses to affirm, and which to reverse. The model makes two predictions about which cases should be reversed and which affirmed. One prediction receives moderate support, the other receives no support. I conclude the chapter with a discussion of other explanations for the patterns observed in the Supreme Court’s dispositional choices.
Chapter 2

Theory

This chapter presents a formal model in which a high court learns about doctrine by aggregating the decisions of multiple lower courts. Although the high court can review only one case, it can see the results of many cases. Allowing the high court to learn from a group of lower courts yields a nuanced relationship between rules and dispositions that is substantively resonant, and leads directly to the conclusion that the high court’s review decisions hinge on estimates of which cases will be most informative to review. The model provides novel and persuasive rationalizations for a number of well-known stylized facts about the judicial hierarchy, including why the Supreme Court finds it so important to review conflict between lower courts, and why the Supreme Court often reviews and reverses the decisions of reliable lower courts. In so doing, it deepens our understanding of how the judicial hierarchy works together to learn and create law.

Before describing the formal structure of the game, I briefly discuss the game’s structure and intuition. In the model, the Supreme Court supervises two lower courts. Each of the three courts wishes to choose the best doctrine to fit a new legal question. For example, when police conduct warrantless searches in motorhomes, the courts must decide whether
the appropriate doctrine comes from searches of houses or searches of cars (see California v. Carney and Friedman (2006)). In such cases, the justices seek to learn facts about the world that make one doctrine or another more appropriate. Often, these are best understood as social-scientific facts. For example, in the case of the motorhome search, the justices sought to understand how owners relate to their motorhomes, referencing Motor Home and RV Lifestyle magazines and studying the interior of the motorhome for signs that it functioned as a living space (California v. Carney, Justice Stevens, dissenting.) In the model, the two lower court judges hear lawyers' arguments for both sides of the dispute, then decide their cases. The Supreme Court sees the decisions the lower court judges make, but does not hear the arguments that led to those decisions. Still, although the lawyers' arguments are unknown, the Supreme Court can draw simple inferences about them from the judges’ choices.

This is the crux of the model’s intuition: even though the Supreme Court can review only one case, it can observe the results of many cases. Therefore, learning can take place before certiorari is granted—much more learning than is usually assumed or considered in the extant literature. This learning allows the Supreme Court to make an informed choice about which case to review. In particular, the justices can distinguish when a lower court judge has made a moderate decision and when his decision is immoderate. The justices can also make reasonably strong deductions about the arguments that led to each. In some instances, it is obvious what arguments must have been presented—an unbiased judge only makes an extreme decision if one party’s evidence was much stronger than the other’s. Other decisions are ambiguous—moderate decisions can arise either because strong arguments were presented for both liberal and conservative positions or because both sides’ arguments were weak. The Supreme Court can let the lower courts’ decisions stand, or it can choose to review one of the lower courts’ decisions, at some cost, before announcing the final doctrine. Reviewing
the ambiguous case will always be more informative for choosing optimal doctrine; therefore, the ambiguous decision is more likely to be reviewed. After review, some information allows the Supreme Court to make dispositive rulings while other information is only suggestive. As a result, the Supreme Court may either reverse or affirm after review. The sections that follow present equilibria describing what choices lower court judges will make, which cases the Supreme Court will review, and what the Supreme Court will do upon review.

2.1 The Model

The model uses the architecture of the model in Dewatripont and Tirole (1999). In this model, there are three players in the game: two unitary lower courts, $LC_I$ and $LC_{II}$, and one unitary Supreme Court. For simplicity, I refer to the lower courts as “judges” and occasionally refer to a lower court judge as “he.” I refer to the Supreme Court as “it.” The goal is to choose one of three doctrines—$A$, $M$, or $B$—to apply. One doctrine will be best in the sense that it best suits the state of the world. Judges will always prefer the best doctrine, but because the area of law is relatively new, they do not know which one that is. I assume that there are two unknown state variables, $\theta_A$ and $\theta_B$, that together determine the state of the world. Payoffs to the courts depend on the conjunction of both variables and the choice of doctrine. A sufficient summary of the state is $\theta = \theta_A + \theta_B$. It is common knowledge that:

$$
\theta_A = \begin{cases} 
0 & \text{with prob. } 1 - \alpha \\
-1 & \text{with prob. } \alpha 
\end{cases}
$$

$$
\theta_B = \begin{cases} 
0 & \text{with prob. } 1 - \alpha \\
1 & \text{with prob. } \alpha 
\end{cases}
$$

Thus

$$
\theta = \begin{cases} 
-1 & \text{with prob. } \alpha(1 - \alpha) \\
0 & \text{with prob. } 1 - 2\alpha + 2\alpha^2 \\
1 & \text{with prob. } \alpha(1 - \alpha) 
\end{cases}
$$

For every state of the world there is an associated doctrine: $A$ if $\theta = -1$, $M$ if $\theta = 0$, and $B$ if $\theta = 1$. $A$, $M$, and $B$ represent existing doctrines or approaches, which might be thought
of as liberal, moderate, and conservative policies, respectively. The Court can be thought to be extending these by deciding which is most applicable for a new fact pattern. An example of this is sex discrimination law, in which judges struggled with the choice between rational basis review and strict scrutiny and ultimately created the doctrine of intermediate scrutiny. Of course, most cases at the Courts of Appeals are simple applications of existing law; this model focuses on the subset of difficult, law-creating cases, either “gap filling” or cases of first impression in which multiple doctrines could plausibly be applied.

The game proceeds in five stages, two in the lower courts and three in the Supreme Court. First, lawyers present evidence to the lower courts about the value of $\theta$. Second, each lower court judge makes a decision based on the evidence he sees. Third, the Supreme Court sees the lower court judges’ decisions, but does not see the evidence that led to those decisions. It uses this information to update its beliefs about $\theta$. Fourth, the Supreme Court decides whether, and which, case to review—the Supreme Court can review at most one case. If the Supreme Court reviews, it learns the messages that lower court saw, then makes its decision—whether to affirm or reverse the decision it reviewed and which doctrine to choose. I discuss each of these steps in detail below; the game is summarized in Figure 1.

2.1.1 Decision making in the lower courts

Simultaneously, the lower courts each hear a case. Both cases depend on the value of $\theta$, which is common across both courts. In this sense, arguments are interpreted in the same manner as by Che and Kartik (2009): “The signal could take the form of scientific evidence obtainable by conducting an experiment, witnesses or documents locatable by investigation, a mathematical proof, or a convincing insight that can reveal something about the state.” Legally, they are appropriately interpreted as legislative facts (which are often solved by expertise and may pertain to many cases), as opposed to adjudicative facts (which pertain to a particular party); see Davis (1942). To decide the case correctly, the judges must learn
the value of $\theta$; because $\theta$ cannot be observed directly, this means learning about $\theta_A$ and $\theta_B$. Two lawyers—one in each lower court—search for evidence about $\theta_A$. Their searches are independent. The same is true for $\theta_B$: two lawyers, one in each lower court, independently search for evidence. Each lawyer then privately presents the results of his search to his lower court judge. $m_A$ denotes the messages of the lawyers for $\theta_A$; $m_B$ denotes the messages of the lawyers for $\theta_B$. Each message takes on one of two values: for $i \in \{A,B\}$ a lawyer either finds and presents hard evidence $|m_i| = 1$ to the judge, or does not find any conclusive evidence and so presents $m_i = 0$.

I discuss the game as if lawyers are presenting evidence to the court, but abstract away from strategic advocacy by the lawyers—I assume that incentives are such that a lawyer presents any evidence he finds and assume lawyers cannot fabricate evidence, so lawyers’ messages are always truthful. The incentives that maintain this condition are the focus of Dewatripont and Tirole (1999). From their results it is possible to deduce that promising the lawyers sufficiently high wages can always satisfy this condition, so long as the lawyers care only about winning their own case.
If $\theta_i = 0$, both lawyers are unable to find any hard evidence and send messages $m_i = 0$. If $|\theta_i| = 1$, each lawyer finds hard evidence of this with probability $q$. When he finds evidence that $|\theta_i| = 1$, a lawyer sends message $|m_i| = 1$. Even if $|\theta_i| = 1$, however, a lawyer may fail to find evidence of this fact. That is, the lawyer may not find evidence that exists, even when he is searching for it. This happens with probability $1 - q$. In this instance, the lawyer sends message $m_i = 0$ even though $|\theta_i| = 1$. Therefore, when a lawyer for $\theta_A$ presents no hard evidence, this suggests $\theta_A = 0$, as it is also possible that $\theta_A = -1$ but the lawyer did not find the evidence. In contrast, a message of $m_A = -1$ proves $\theta_A = -1$. Thus, presenting evidence perfectly reveals the state of the world, but failing to present evidence is merely suggestive.

Notice also that if $\theta_i = 0$ both lawyers will send $m_i = 0$, but if $|\theta_i| = 1$ the lawyers may send different messages if one’s search is successful and the other’s is not. However, each lower court judge observes only his own lawyers’ messages—he cannot learn what the other lower court did or what messages the other lower court received.

Thus, a lower court judge observes one of four possible message pairs—$(0, 0)$, $(0, 1)$, $(-1, 0)$, or $(-1, 1)$. After observing one of these pairs, each lower court judge makes an inference about the value of $\theta$, which incorporates the primitive probability that $|\theta_i| = 1$, $\alpha$; and the conditional probability that a lawyer’s search is successful, $q$. After establishing a posterior belief about the value of $\theta$, each lower court judge makes a decision, $A$, $M$, or $B$, to correspond to his belief. Denote this ruling $D$: $D_I$ for $LC_I$ and $D_{II}$ for $LC_{II}$.

### 2.1.2 Learning and decision making at the Supreme Court

Both cases are then automatically appealed to the Supreme Court. The Supreme Court can review either one of the lower courts’ decisions, or neither, but not both. The Supreme Court may consolidate cases and hear them together. I ignore this option to maintain a focus on the Supreme Court’s choices when it does not have the resources to read every lower court’s case on a particular question. Section 2.2.3 considers the implications of allowing the Supreme Court to review both lower courts.
Court sees both lower courts’ rulings but does not directly observe the evidence the judges saw. In terms of verisimilitude, this is a reasonable stylization of the appeals process: lower courts’ rulings are presented in the briefs petitioning for review, while lawyers’ arguments are only submitted if the Supreme Court chooses to review the case. After seeing the lower courts’ rulings, the Supreme Court updates its beliefs about $\theta$ and decides whether to review either of the lower courts’ decisions.

If the Supreme Court chooses not to review a case, the lower courts’ decisions stand and the game ends. If the Supreme Court does choose to review a case it learns the messages that judge saw, but it must also pay a cost of review $c$. This parameter encompasses the opportunity cost of reviewing said case (instead of a case on a different matter) and the time and resources expended reading briefs and hearing arguments. Once the Supreme Court has heard the arguments presented in that case, it uses this information to further update its beliefs about $\theta$. (Note that the messages are preserved perfectly between the Courts of Appeals stage and the Supreme Court stage; there is no additional information collection between the stages.) Based on its estimates of $\theta$, the Supreme Court then chooses a disposition and a doctrine. The disposition, to reverse or affirm, pertains only to the case it is reviewing. The doctrine, $A$, $M$, or $B$, is a universally binding precedent that can effectively reverse or affirm the decision not reviewed. That is, issuing a universally binding doctrine changes the outcome of all cases, even those the Supreme Court did not review. Like the lower court judges, the Supreme Court chooses the doctrine that matches its beliefs about $\theta$. Its decision to reverse or affirm the lower court’s ruling follows immediately from this doctrinal choice—it affirms their decision if it agrees the decision is appropriate based on its own estimate of $\theta$. Of course, the Supreme Court’s estimate of $\theta$ may be different from the lower court’s estimate, for although neither can see the arguments presented in the unreviewed lower court, the Supreme Court’s beliefs are also based on the additional
Types of Judicial Preferences

<table>
<thead>
<tr>
<th></th>
<th>Unbiased</th>
<th>Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>-L 0 -L</td>
<td>0 -L -L</td>
</tr>
<tr>
<td>B</td>
<td>-1 -L 0</td>
<td>-1 -1</td>
</tr>
</tbody>
</table>

Table 2.1: Judges’ preferences over doctrine, A, M, or B, conditional on the state of the world \( \theta \), -1, 0, or 1. All judges get the maximum utility, 0, from choosing the right doctrine—A when \( \theta = -1 \), M when \( \theta = 0 \), and B when \( \theta = 1 \). Mistakes cost \(-1\) or \(L\), where \(0 < L < 1\). The left panel shows the preferences of unbiased judges who lose more utility from large mistakes than small ones, but have symmetric preferences otherwise. The right panel shows judges who are biased against B, so that wrongly choosing B is more costly than wrongly choosing A.

information provided by the unreviewed lower court’s decision, which the reviewed lower court cannot see.

2.1.3 Preferences and beliefs

Before the game begins, each judge believes \(pr(\theta_A = -1) = pr(\theta_B = 1) = \alpha\), and believes that if \(|\theta_i| = 1\) a search is successful with probability \(q\), that is,

\[
pr(m_A = -1|\theta_A = -1) = pr(m_B = 1|\theta_B = 1) = q.
\]

After seeing messages from the lawyers, a lower court judge is able to update his beliefs about \(\theta\). Lower court judges update their beliefs based only on their own advocates’ messages. Thus, after hearing arguments, \(LC_I\)’s beliefs about \(\theta\) are a function of \((\alpha, q, m_{A_{II}}, m_{B_{II}})\) and \(LC_{II}\)’s beliefs about \(\theta\) are a function of \((\alpha, q, m_{A_{II}}, m_{B_{II}})\). The Supreme Court is able to update its beliefs about \(\theta\) based on both lower courts’ decisions. After seeing the lower courts’ decisions, the Supreme Court’s beliefs about \(\theta\) are a function of \((\alpha, q, D_{I}, D_{II})\). If the Supreme Court chooses to review one of the lower courts’ decisions, it learns the evidence that lower court received. This allows it to update its beliefs again. Then, its beliefs are a function of \((\alpha, q, m_{A_{I}}, m_{B_{II}}, D_{II})\) if it reviews \(LC_I\) or \((\alpha, q, D_{I}, m_{A_{II}}, m_{B_{II}})\) if it reviews \(LC_{II}\).
All judges agree on the best doctrine when they know the value of $\theta$ with certainty—$A$ is best when $\theta = -1$, $M$ is best when $\theta = 0$, and $B$ is best when $\theta = 1$. But judges may differ in their views of the costs for certain types of mistakes, so when there is uncertainty about the value of $\theta$ they may disagree about which doctrine to choose. Consider a suit brought by an injured car owner against the manufacturer, where the judge must decide if the manufacturer’s safety efforts met a standard of care. If the manufacturer is indeed liable for an injury that he should have prevented, all judges agree he should be penalized. But if there is uncertainty about whether or not he is liable then the judges might disagree as to the best outcome: some may believe the manufacturer should not be overburdened with requirements based on inconclusive claims of liability, while others might believe that protecting the consumer should take precedence. This is formalized by letting some judges suffer more from choosing $A$ than $B$ when the correct decision is $M$. Furthermore, under certain conditions, a judge’s fear under uncertainty can be so extreme that one lawyer could never provide enough evidence to convince him to choose a particular result. For example, a judge biased in favor of drivers might only be willing to choose a low standard of care if all evidence suggests manufacturers are never liable, so that one lawyer could never present enough evidence in one case to convince him of such.

Because all judges agree what they should do if the facts are clear, a judge who chooses the doctrine that corresponds to the state of the world always gets utility 0. If the doctrine he chooses is wrong, he incurs some cost; these costs vary across judges and doctrines. The panels of Table 2.1 show different arrangements of these costs. Consider the lefthand panel. In that panel, a judge loses 1 if he chooses $A$ when $\theta = 1$ or $B$ when $\theta = -1$. This is a bad mistake, where there is a large mismatch between doctrine and the state of the world. If he makes a smaller mistake—choosing $A$ or $B$ when $\theta = 0$, or $M$ when $\theta = -1$ or 1—the judge loses $L$, where $0 < L < 1$. Thus, if a judge chooses doctrine $M$, for example, his expected
utility is \(-L \cdot pr(\theta = -1) - L \cdot pr(\theta = 1)\). In the righthand panel, the judge is wary of choosing doctrine \(B\). This is formalized by making a small mistake as costly as a large one, so that choosing \(B\) when \(\theta = 0\) costs 1. But choosing \(A\) when \(\theta = 0\) still costs this judge only \(L\). This imbalance captures judicial bias—the judge is willing to choose doctrine \(A\), even if it might be the wrong doctrine, but he is less willing to choose doctrine \(B\), even if it might be the right doctrine. Notice this bias pushes a judge toward moderation—rather than leading lower court judges to choose \(B\) when evidence suggests they choose \(M\), this operationalization makes biased judges unwilling to risk movements away from moderate doctrine. Such a conception might mean, for example, that a biased judge fears a slippery slope, or that a liberally-biased judge is unwilling to extend conservative doctrine in a case that is plausibly distinguished from conservative precedent.

Finally, note that each judge’s utility is affected only by his ruling and the true state of the world, not the rulings of others. In other words, lower court judges care only about resolving the dispute correctly based on the evidence they see, without concern for future doctrine or response from the Supreme Court.

### 2.2 Optimally Learning from Agents’ Decisions

To understand how the Supreme Court learns from inferior courts, I consider two arrangements of preferences. First, as a baseline, I consider lower courts whose preferences are identical to one another and to the Supreme Court. Presented with the same information, every judge in this version of the game would make the same decision. The equilibrium from this game is presented in Section 2.2.1. I then consider a scenario where the Supreme Court supervises one ideological ally and one judge who is biased. Section 2.2.2 presents the

---

3I choose to assume lower court judges do not fear reversal for two reasons. First, even if Courts of Appeals judges fear reversal, this is not likely to come into play in cases of first impression. Second, the assumption highlights the challenge of learning from agents who are purely self-interested.
equilibrium under these conditions. All proofs appear in the appendix.

2.2.1 Supervising two unbiased judges

I begin with the choices of the lower court judges, who attempt to resolve cases based on the evidence the lawyers present. A lower court judge wants to choose the outcome that best corresponds to the state of the world. He learns the probability of each state, \( \theta \in \{-1, 0, 1\} \), from the lawyers’ arguments. Recall that before he sees the results of a lawyer’s search for evidence, the judge’s prior belief that \( \theta_A = -1 \) is \( \alpha \). He also believes the probability that \( \theta_B = 1 \) is \( \alpha \). Suppose a lawyer’s search is unsuccessful, so the judge receives a message \( m_i = 0 \). Define the judge’s posterior belief \( \text{pr}(\theta_A = -1|m_A = 0) \equiv \hat{\alpha} = \frac{\alpha - \alpha q}{1 - \alpha q} \) (and likewise \( \text{pr}(\theta_B = 1|m_B = 0) \equiv \hat{\alpha} \)). This posterior belief encapsulates the chances that \( |\theta_i| = 1 \) and the lawyer was simply unsuccessful in proving this. I place restrictions on \( \hat{\alpha} \) so that after observing \( m_i = 0 \), the lower court judge is more inclined to believe that \( \theta_i = 0 \) than \( |\theta_i| = 1 \). This condition is \( \hat{\alpha} < 1/2 \). Because of this assumption, if a lower court judge receives a message pair of \((-1, 0)\), he believes it is more likely that \( \theta = -1 \) than that \( \theta = 0 \). (Since he knows \( \theta = \theta_A + \theta_B, \theta_A = -1, \text{ and } \theta_B \in \{0, 1\}, \) he knows \( \theta \neq 1 \)) In other words, after seeing \((-1, 0)\) he believes it is more likely that \( A \) is the best doctrine than that \( M \) is. But he is not sure—it is possible that \( \theta_B = 1 \) and the lawyer failed to find evidence of this, in which case \( \theta = 0 \) and \( M \) is the best doctrine.

Since judges are unbiased in this version of the game, a lower court judge suffers equal utility loss if he chooses \( A \) when he should have chosen \( M \) or if he chooses \( M \) when he should have chosen \( A \) (and likewise for \( B \)). As a result, after hearing arguments, the lower court judge decides which state is most likely given the probabilities described above, and chooses the associated doctrine. Messages \((-1, 0)\) and \( \hat{\alpha} < 1/2 \) imply \( \theta = -1 \) is most likely; therefore a judge who sees \((-1, 0)\) will choose \( A \). The same is true for \((0, 1)\)—this will lead the lower court judge to choose \( B \). If he receives a message pair of \((-1, 1)\), a lower court judge will
choose doctrine $M$, for he knows $\theta = 0$ with certainty. If the lower court judge receives a message pair of $(0, 0)$, there is still a strictly positive probability on all values of $\theta$. If the lower court judge chooses $A$, and $\theta = 1$, he will experience a large loss in utility. Likewise, it will be very costly to choose $B$ if it happens that $\theta = -1$. Choosing $M$ guarantees the lower court will not incur too large a loss, no matter what the value of $\theta$ is. After $(0, 0)$, therefore, the lower court judge will choose doctrine $M$. To summarize, the lower court judge will choose $A$ if and only if he receives messages $(-1, 0)$. Likewise, he will choose $B$ if and only if he receives messages $(0, 1)$. But he will choose $M$ after either $(0, 0)$ or $(-1, 1)$.

This leads to the first stage of Supreme Court inference. If lower courts are behaving optimally, then after some histories the Supreme Court can perfectly infer what messages a judge must have received without reviewing the case. This occurs after a lower court reaches a decision of $A$, in which case the Supreme Court can be sure that lower court must have received messages $(-1, 0)$, or after a lower court makes a decision of $B$, in which case the Supreme Court can be sure that lower court received messages $(0, 1)$. On the other hand, when the Supreme Court observes a decision of $M$, it does not know if it was reached because of messages $(0, 0)$ or $(-1, 1)$. This uncertainty is the primary driving force behind the results that follow: the Supreme Court can only learn the messages prompting a ruling of $M$ by paying a cost, $c$, to review the case. Although the Supreme Court does not directly observe the lawyers’ messages, it has an advantage that the lower courts do not: it sees the results of two cases, and can make an informed decision about whether it is worthwhile to review a case, and if so, which one.

When both lower courts issue the same ruling, $A$, $B$, or $M$, there is nothing to gain from review. Any of these decisions could be wrong, but no review will lead the Supreme Court to learn enough to change the lower courts’ doctrine. If the courts both decide $A$ or $B$, the Supreme Court can perfectly infer the lawyers messages, and because it shares the same
beliefs and preferences as the lower courts, it would rule the same way. If both lower courts
decide M, review will be informative—it will change the Supreme Court’s beliefs about θ—but it will never be outcome-consequential, as the Supreme Court will always choose M. If the Supreme Court observes one lower court choose A and the other choose B, the Supreme Court concludes with certainty that θ = 0 without reviewing either case—but it still must pay c in order to communicate this to the lower courts. Since either case is an equally good vehicle, it randomly chooses one to review. It reverses the decision and announces a doctrine of M. If one court rules either A or B while the other court rules M, the Supreme Court could learn valuable information by reviewing the case that generated the ruling of M. Suppose LC_I has made a decision of A and LC_{II} has made a decision of M. Under these conditions, the Supreme Court can perfectly infer the messages that LC_I saw—they must have been (−1, 0). Based only on the fact that LC_I chose A, the Supreme Court knows for sure that θ_A = −1 and is slightly more confident that θ_B = 0 than before. It uses this information to make an inference about the messages LC_{II} saw, knowing it is more likely that LC_{II} received evidence that θ_A = −1 and less likely that LC_{II} received evidence that θ_B = 1. Then the Supreme Court decides whether to pay c to review LC_{II}’s decision. If it discovers LC_{II}’s decision was generated by messages of (−1, 1), the Supreme Court learns with certainty that a decision of M is correct. If LC_{II}’s decision was generated by messages of (0, 0) the Supreme Court is much more inclined to believe the appropriate doctrine is A than M, but it still does not know this with certainty and so finds it less beneficial to reverse the decision than otherwise. It will review LC_{II}’s decision if either the probability of learning (−1, 1), or the costs from an incorrect decision, are sufficiently high. These beliefs and actions describe the equilibrium in the game with homogeneous agents.

Proposition 1 (Equilibrium with homogeneous agents) In the game with homogeneous agents, the following occurs in the unique equilibrium.
Each lower court chooses:

\[
\begin{cases}
A & \text{iff he receives messages } (-1, 0) \\
B & \text{iff he receives messages } (0, 1) \\
M & \text{iff he receives messages } (0, 0) \text{ or } (-1, 1)
\end{cases}
\]

After seeing the lower courts’ decisions, the Supreme Court does the following.

- **If the lower courts chose** \((A, A), (B, B), \text{ or } (M, M)\), **the Supreme Court does not review a case.**

- **If the lower courts chose** \((A, M)\), **the Supreme Court reviews** \(LC_{11}\) if

\[
c < L \left( 1 - \frac{\alpha(1 - q)^2}{\alpha(1 - q)^2 + q^2 + 1} \right)
\]

otherwise it does not review either case.

  - If it discovers \(M\) was generated by messages \((-1, 1)\), it determines \(\theta = 0\), affirms the decision of \(M\), and issues universal precedent \(M\).
  
  - If it discovers \(M\) was generated by \((0, 0)\), it believes \(\theta = -1\) with \(p > 1/2\), reverses the decision of \(M\), and issues universal precedent \(A\).

  - Parallel equilibrium strategies hold for \((M, A), (B, M), \text{ and } (M, B)\).

- **If the lower courts chose** \((A, B)\) or \((B, A)\), the Supreme Court determines \(\theta = 0\). If \(c < 2L\), it reviews a case (either case), reverses the decision, and issues universal precedent \(M\).

The most notable result in this equilibrium is the tendency to review moderate decisions. Given that the Supreme Court can afford to review only one case, it will be most likely to review decisions of \(M\). When costs are low and one lower court has made a decision of \(M\) while the other has not, the Supreme Court is more likely to review the decision of \(M\) than the other. In fact, when costs are low, a decision of \(M\) is always reviewed unless both lower courts reach a decision of \(M\). This is because reviewing a decision of \(M\) is always informative, and outcome-consequential unless both lower courts make that decision. (Recall that reviewing after both lower courts choose \(M\) may improve the Supreme Court’s certainty in its decision, but will still always lead it to choose doctrine \(M\).) In contrast, a decision of
A is never informative to review. Thus, decisions of A are only reviewed if the other lower court makes a decision of B; even then, the Supreme Court may randomly choose to review the other case.

However, as the cost of review rises, decisions of M become less likely to be reviewed. This is because observing simultaneous decisions of A and B guarantees maximum utility upon review, while reviewing a decision of M is less beneficial in expectation. Thus, when costs are moderately high, the Supreme Court is more likely to review extreme conflict (where one lower court chooses A and the other chooses B) than moderate conflict (where one lower court chooses M and the other does not). Furthermore, in this game the Supreme Court will never review if there is no conflict. Black and Owens (2009) find that the Supreme Court is particularly likely to review conflicts that are “neither shallow nor tolerable.” The model provides a persuasive theoretical explanation for why conflicts between the lower courts are so often reviewed, including a justification for why the extremity of conflict matters.

Even though all judges have the same preferences, the Supreme Court still reviews and reverses lower courts’ decisions. In fact, if costs are moderate, all of the Supreme Court’s decisions will be reversals, even though the lower courts are perfectly faithful agents. Recall that the purpose of reviewing a moderate decision is informational, not reversal. As a result, it is not always true that the Supreme Court is more likely to reverse decisions than affirm them—when c is low enough that the Supreme Court is willing to review decisions of M, and α and q are jointly relatively small, then the likelihood of reversal exceeds the likelihood of affirmance. Furthermore, given that the Supreme Court reverses one and only one lower court, it is more likely to be the court it reviews when αq < 1/2.

2.2.2 Bias in the lower courts

Next, to investigate how ideological heterogeneity affects learning, I consider a scenario in which the Supreme Court is supervising one lower court who shares its unbiased ideological
preferences (LCI) and one lower court who is biased against outcome B (LCII). LCII prefers to choose B if \( \theta = 1 \), but if \( \theta = 0 \) he incurs a large loss from choosing B. LCII will therefore only choose B if he is very sure \( \theta = 1 \). To consider the full effects of this bias, I put an additional condition on \( \hat{\alpha} \) so that after seeing \((0,1)\) LCII is not sure enough that \( \theta = 1 \) to be willing to choose B. This condition is \( L < \frac{\hat{\alpha}}{1 - \hat{\alpha}} \) \(^4\). Because of this assumption, LCII’s loss from a ruling of B when \( \theta = 0 \) is larger than that from a ruling of M when \( \theta = 1 \). Thus, LCII chooses M after seeing \((0,1)\).

LCII’s bias means the Supreme Court cannot learn as much about \( \theta \) before deciding whether to review. Now, the only time the Supreme Court chooses not to intervene is when both lower courts decide A. In every other situation, the Supreme Court will review one of the lower courts’ decisions, as long as its cost of review is sufficiently low.

**Proposition 2 (Equilibrium with Heterogeneous Agents)** In the game with heterogeneous agents and a biased lower court, the following occurs in the unique equilibrium.

*Lower Court I chooses:*  
\[
\begin{align*}
A & \quad \text{iff he receives messages } (−1,0) \\
B & \quad \text{iff he receives messages } (0,1) \\
M & \quad \text{if he receives messages } (0,0) \text{ or } (−1,1)
\end{align*}
\]

*Lower Court II chooses:*  
\[
\begin{align*}
A & \quad \text{iff he receives messages } (−1,0) \\
M & \quad \text{if he receives messages } (0,1), (0,0) \text{ or } (−1,1)
\end{align*}
\]

After seeing the lower courts’ decisions, the Supreme Court does the following.

- **If the lower courts chose** \((A,A)\), the Supreme Court does not review a case.
- **If the lower courts chose** \((A,M)\), then the Supreme Court reviews LCII if
  \[
  c < L \left( 1 - 2 \frac{\alpha(1 - q)^2}{\alpha(1 - q)^2 + \alpha q^2} \frac{\alpha(1 - q)^2}{\alpha(1 - q)^2 + 1 - \alpha} \right),
  \]
  otherwise it does not review a case. If it reviews and discovers M was generated by messages
  \[
  - (0,0), \text{ then the Supreme Court reverses LCII’s decision and issues doctrine A.}
  \]

\(^4\)It would also be sufficient to increase the loss from choosing B when \( \theta = 0 \) to an amount greater than 1. I choose to manipulate \( \hat{\alpha} \) instead for algebraic simplicity.
– $(−1, 1)$, then the Supreme Court affirms $LC_{II}$’s decision and issues doctrine $M$.
– $(0, 1)$, then the Supreme Court affirms $LC_{II}$’s decision and issues doctrine $M$.

• If the lower courts chose $(M, A)$, then the Supreme Court reviews $LC_{I}$ if

\[ c < L \left( 1 - \frac{\alpha(1-q)^2}{\alpha(1-q)^2 + q^2 + 1} \right), \]

otherwise it does not review a case. If it reviews and discovers $M$ was generated by messages

– $(−1, 1)$, then the Supreme Court affirms $LC_{I}$’s decision and issues doctrine $M$.
– $(0, 0)$, then the Supreme Court reverses $LC_{I}$’s decision and issues doctrine $A$.

• If the lower courts chose $(M, M)$, then the Supreme Court reviews $LC_{II}$ if $\alpha < 2q(1-q)$ and $c < c^*(L, \alpha, q)$, otherwise it does not review a case. If it reviews and discovers $M$ was generated by messages

– $(0, 0)$ or $(−1, 1)$, then it affirms the decision and issues doctrine $M$.
– $(0, 1)$, then it reverses $LC_{II}$’s decision and issues doctrine $B$.

• If the lower courts chose $(B, A)$ then the Supreme Court takes either case if $c < 2L$, reverses the case, and issues doctrine $M$. If $c \geq 2L$, it does not review.

• If the lower courts chose $(B, M)$ then the Supreme Court reviews $LC_{II}$ if

\[ c < L \frac{1-\alpha}{\alpha(1-q)^2 + 1 - \alpha}, \]

otherwise it does not review a case. If it reviews and discovers $M$ was generated by messages

– $(0, 0)$, then the Supreme Court reverses $LC_{II}$ and issues doctrine $B$.
– $(−1, 1)$, then the Supreme Court affirms $LC_{II}$ and issues doctrine $M$.
– $(0, 1)$, then the Supreme Court reverses $LC_{II}$ and issues doctrine $B$.

This equilibrium differs from the equilibrium with two unbiased courts in two important ways. First, recall that under ideological homogeneity, the Supreme Court grants *certiorari* only if there is conflict in the lower courts. Even though the Supreme Court would learn the fact pattern that led to one of the courts’ choices, review without conflict would never

\[ ^5 \text{See proofs for formula for } c^*. \]
be outcome-consequential. This result does not hold when one of the lower courts is biased; now, the Supreme Court does review after the courts reach the same conclusion. This is because the lower courts might reach the same conclusion for different reasons, and that possibility merits the Supreme Court’s attention.

Second, as a result of this, the Supreme Court may find that all cases on some matter were decided correctly. Resolving conflict requires the Supreme Court to reverse at least one of the lower courts’ decisions, so without biased lower courts an affirmanse is always accompanied by an implicit reversal of the other court’s decision. Now, however, the Supreme Court may affirm two decisions that are similar on their face (even though they were generated by different arguments). A closely related result is that the probability of an affirmanse is higher when one lower court is biased. In all histories in which the Supreme Court affirms with two unbiased lower courts, it also affirms when one lower court is biased. With a biased lower court, however, it also (1) affirms cases it would have not have reviewed and (2) affirms cases it would have reversed, had both lower courts been unbiased.

That the Supreme Court affirms cases it would not have reviewed highlights an important point: the probability of review is higher with a biased lower court than without. When one court is biased, the Supreme Court grants certiorari to all cases it would review under homogeneity as well as to additional cases. Furthermore, most of this additional review falls on the biased agent, who now chooses $M$ and earns review when his messages are $(0, 1)$. As a result, the Supreme Court is more likely to review the biased lower court than its ideological ally. This result is similar to previous models of the judicial hierarchy (like Cameron, Segal and Songer (2000)), but the result is more subtle: occasionally the Supreme Court will prefer reviewing its ideological ally to reviewing the biased lower court. For example, when $LC_I$ has made a decision of $M$ and $LC_{II}$ has made a decision of $A$, the Supreme Court will prefer to review its ally over the biased lower court. Also, the intuition behind this result...
differs from that of Cameron, Segal and Songer (2000): the Supreme Court does not review
\( LC_{I_1} \) because he might have cheated but rather because his lawyers are silenced unless the
Supreme Court reviews his decisions.

The second of these effects—that, *conditional* on review, the likelihood of an affirmance
is higher with a biased lower court—arises because biased decisions are sometimes affirmed
in equilibrium. This occurs when the unbiased lower court chooses \( A \), and the biased lower
court chooses \( M \) despite receiving messages which would lead an unbiased judge to choose
\( B \). Together, these messages guarantee that the appropriate doctrine is \( M \). \( LC_I \)'s decision
of \( A \) implies \( \theta_A = -1 \), and the messages from \( LC_{I_1} \)'s lawyers—\((0,1)\)—imply \( \theta_B = 1 \). Thus,
after seeing \((A,M)\) and reviewing \( LC_{I_1} \)'s decision, the Supreme Court knows \( \theta = -1 + 1 = 0 \). Therefore, even though \( LC_{I_1} \) behaved contrary to how the Supreme Court would have
wanted, the Supreme Court upholds his decision.

Although counter-intuitive, it is important to note that this does, in fact, occur. At
times, the Supreme Court will review a decision and, although it will disagree with the
*reasoning* espoused by the lower court, it will affirm its decision on the merits. For example,
in *U.S. v. Tinklenberg* (131 S.Ct. 2007), Justice Breyer wrote the following in the opinion
of the Court: “We disagree with the Sixth Circuit’s interpretation of both subparagraph
(D) and subparagraph (F), and now hold that its interpretations of those two provisions
are mistaken. Nonetheless the conclusions the court drew from those two interpretations in
relevant part cancel each other out such that the court’s ultimate conclusion ... is correct.
Therefore, the Sixth Circuit’s judgment ordering dismissal of the indictment on remand is
affirmed.”

### 2.2.3 Case consolidation

When the Supreme Court is confronted with multiple petitions for *certiorari* that are
closely related, it has the option of consolidating the appeals and hearing them as one case.
Although this occurs frequently, scant empirical and theoretical attention has been paid to when, why, and how the decision to consolidate is made. This model provides a potential avenue for early exploration of such a question. Consider, for example, a Supreme Court that can pay $2c$ to review both decisions, or $c$ to review either one. The Supreme Court will never choose to review both cases. (Even if the second case cost only some small quantity $\epsilon$ to review, the Supreme Court would still not review both.)

Recall that whenever a lower court makes a decision of $A$ or $B$, the Supreme Court cannot learn anything by reviewing the decision. Recall also that after the lower courts have made decisions $(A, B)$, the Supreme Court is able to reverse and issue a doctrine of $M$ by reviewing only one lower court. Therefore, the Supreme Court has no incentive to review both cases unless both lower courts have made a decision of $M$. Notice then that when both homogeneous lower courts have made a decision of $M$, review is not outcome-consequential—the Court will continue to choose $M$ no matter what it learns. Therefore, it has no incentive to review both cases. Only when heterogeneous lower courts both make decisions of $M$ might the Supreme Court choose to review both cases.

### 2.3 Discussion and Conclusion

Law is not static. As a society changes, new problems appear, new classes of disputes arise, and doctrine must adapt to govern their resolution. This chapter aims to understand the process by which the Court extends doctrine to adjudicate these new cases. In so doing, it joins the literature advancing a perspective on the judicial hierarchy as a learning organization (Kornhauser 1989, Cooter, Kornhauser and Lane 1979, Baker and Mezzetti 2012, Niblett 2013, Clark and Kastellec 2013). This approach contrasts with the dominant mode

---

6According to the Supreme Court Database, in the 2010 term, the Supreme Court issued 85 written opinions, disposing of 95 cases; in the 2011 term, the Supreme Court issued 77 written opinions, disposing of 88 cases. Thus, about 10-12% of the disputes addressed by the Supreme Court were consolidated.
of understanding the hierarchy over the last decade, the disciplinary or hierarchical control perspective. Instead of focusing on how the Supreme Court monitors the resolution of existing disputes, this new literature focuses on understanding the process by which the Supreme Court learns how to extend, develop, and adapt existing doctrine to fit new questions. Outside of the judicial literature, hierarchy is known to create unique specialization schemes: the Supreme Court may specialize in answering difficult questions while relying on agents to answer easier ones [Garicano and Zandt 2013]. This model, however, sees the Supreme Court’s specialization as being in *supervision* rather than in the resolution of similar, but more difficult, questions.

The model demonstrates that the Courts of Appeals serve as laboratories of law: the Supreme Court watches their decisions to learn how best to extend doctrine. Before establishing a rule to govern future lower courts’ decisions, the Supreme Court learns which rule will be best by considering lower courts’ decisions in previous cases. This requires analyzing multiple lower courts’ decisions in concert, using one to gain leverage on the implications of another. The Supreme Court is then able to make informed decisions about which cases to review, and upon review is able to make informed extrapolations from the case at hand. Based on this theory, the model identifies which cases the Supreme Court will choose to review and what doctrine it will support.

Some of the predictions arising from the model are surprising. Because ambiguous cases are more informative, and because extreme decisions are never ambiguous, the Supreme Court should be more likely to review *moderate* decisions than extreme decisions. The empirical results show the Supreme Court does indeed behave this way, which suggests the Supreme Court’s primary focus is legal development, not doctrinal discipline. The Court does not appear to use its position to punish extreme decisions with which it disagrees, as previous literature has suggested.
The model also offers theoretical explanations for a number of stylized facts, including why the Supreme Court focuses on resolving conflicts between lower courts, why the Supreme Court is more likely to review ideologically distant lower courts, and why, despite this propensity, the Supreme Court often reviews and reverses the decisions of its allies. Concurrently, however, the model challenges existing explanations for other empirical patterns. For example, because prior literature has focused on discipline in the hierarchy, many have understood Courts of Appeals judges’ dissenting opinions to be signals of non-compliance (Epstein, Landes and Posner 2011; Kastellec 2007; Beim, Hirsch and Kastellec 2012; though see Hettinger, Lindquist and Martinek 2004). The learning perspective suggests dissents may also be pieces of evidence—perhaps a judge can choose to search as if he were an advocate and write a dissenting opinion that presents the evidence he finds.\footnote{This possibility—where ideological extremism motivates judges to search for information when their colleagues do not—is considered in Spitzer and Talley (2012). Such a model might also bear some resemblance to Gailmard and Patty (2013), in which an agent has observed the results of one search and may choose whether to investigate again.}

Beyond the judicial application, the addition of a second lower court adds the conceptual concerns of consistency and choice of review to the learning dynamics considered in Dewatripont and Tirole (1999). By studying iterative hierarchical learning—how lower court judges learn from lawyers and how higher courts learn from lower courts’ decisions—the model contributes to a broader literature on information-gathering and optimal experimentation in hierarchical organizations. Three extensions to the theoretical model stand out as particularly relevant for further exploration of these questions. First, what would change if the lower courts cared about the final doctrine articulated by the Supreme Court, in addition to caring about the dispositions of their own cases? My model assumes lower court judges’ preferences are myopic; therefore, lower court judges resolve cases based only on the evidence they see, with no eye toward policy-making. It is interesting to consider how...
the model would change if lower court judges feared reversal or wished to aid the Supreme Court in its law-creation pursuits. Second, the model assumes one specific form of judicial bias—namely, it does not consider scenarios in which the Supreme Court itself is biased, or ones in which multiple lower courts are biased. These scenarios would enrich the results of this model, and further extend its applicability. Third, this model ends once the Supreme Court chooses a doctrine. In practice, the Supreme Court monitors the application of its chosen doctrine. Such an extension would re-introduce the disciplinary dynamics that have characterized previous work on the judicial hierarchy; as such, it could integrate established results on optimal monitoring with new results on law creation through experimentation.
Appendix: Proofs

Proof. Ideological Homogeneity.

1. The Lower Courts  When a lower court judge observes $(-1, 1)$ he believes $pr(\theta = 0) = 1$. Since $M$ maximizes utility when $\theta = 0$, the lower court chooses it.

   When a lower court judge observes $(0, 0)$ he believes $pr(\theta = -1) = pr(\theta = 1) = \frac{(1-q)\alpha}{1+\alpha-2qa}$. He believes $pr(\theta = 0) = \frac{1-\alpha}{1+\alpha-2qa}$. Therefore, his expected utility from choosing $A$ or $B$ is $-L\frac{\alpha(1-q)}{1+\alpha-2qa} - L\frac{\alpha(1-q)}{1+\alpha-2qa}$. His expected utility from choosing $M$ is $-L\frac{\alpha(1-q)}{1+\alpha-2qa} - L\frac{\alpha(1-q)}{1+\alpha-2qa}$. His expected utility from $M$ is therefore greater:

   $$-L\frac{\alpha(1-q)}{1+\alpha-2qa} - L\frac{\alpha(1-q)}{1+\alpha-2qa} - \left[ -L\frac{\alpha(1-q)}{1+\alpha-2qa} - L\frac{\alpha(1-q)}{1+\alpha-2qa} \right] = -L\alpha(1-q) + \alpha(1-q) - L\alpha(1-q) + L(1-\alpha)$$

   $$= L(1-\alpha) + (1-q)(\alpha - 2L\alpha) > 0.$$ 

   He chooses $M$.

   I place restrictions on $\hat{\alpha}$ so that after observing $(-1, 0)$ the lower court judge is more inclined to believe $\theta = -1$ than $\theta = 0$ and after observing $(0, 1)$ the lower court judge is more inclined to believe $\theta = 1$ than $\theta = 0$. These conditions are $\hat{\alpha} < 1/2$. Thus, lower courts choose $M$ after $(0, 0)$ and $(-1, 1)$; $A$ after $(-1, 0)$, and $B$ after $(0, 1)$.

2. The Supreme Court  Given lower court judges’ strategies, the Supreme Court makes the following inferences after each observed history:

   After $(A, A)$ the Supreme Court knows messages must have been $(-1, 0; -1, 0)$. It can gain no information from taking a case and so does not. The same holds for $(B, B)$.

   After $(M, M)$ the Supreme Court knows messages must have been either $(0, 0; 0, 0)$ or $(-1, 1; 0, 0)$ or $(0, 0; -1, 1)$ or $(-1, 1; -1, 1)$. By taking a case it can gain utility from certainty but it will never reverse. Therefore, it will never review after seeing $(M, M)$. Table 2.2 shows all possible histories and the utility change from review.
State of the world Messages Utility change
\[ \theta = -1 \ (0,0;0,0) \quad -2L \text{ without review, } -2L \text{ with review} \]
\[ \theta = -1 \ (-1,1;0,0) \quad \text{Cannot happen} \]
\[ \theta = -1 \ (-1,1;-1,1) \quad \text{Cannot happen} \]
\[ \theta = 1 \ (0,0;0,0) \quad -2L \text{ without review, } -2L \text{ with review} \]
\[ \theta = 1 \ (-1,1;0,0) \quad \text{Cannot happen} \]
\[ \theta = 1 \ (-1,1;-1,1) \quad \text{Cannot happen} \]
\[ \theta = 0 + 0 \ (0,0;0,0) \quad 0 \text{ without review, 0 with review} \]
\[ \theta = 0 + 0 \ (-1,1;0,0) \quad \text{Cannot happen} \]
\[ \theta = 0 + 0 \ (-1,1;-1,1) \quad \text{Cannot happen} \]
\[ \theta = -1 + 1 \ (0,0;0,0) \quad 0 \text{ without review, 0 with review} \]
\[ \theta = -1 + 1 \ (-1,1;0,0) \quad 0 \text{ without review, 0 with review} \]
\[ \theta = -1 + 1 \ (-1,1;-1,1) \quad 0 \text{ without review, 0 with review} \]

Table 2.2: All possible states of the world that could generate decisions \((M, M)\). The Supreme Court never gains from review.

**After \((A, M)\) or \((M, A)\) or \((B, M)\) or \((M, B)\)** the Supreme Court wishes to learn whether \(\theta = 0\) and the message simply failed to reveal this or whether evidence suggests \(|\theta| = 1\). It chooses to review the \(M\) decision to learn this. It will do so whenever the expected potential benefit of additional information outweighs the cost \(c\) of taking the case. The Supreme Court’s expected utility change after review is as follows:

<table>
<thead>
<tr>
<th>State of the world</th>
<th>Messages</th>
<th>Utility change</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta = 1 ) \ (-1,0;1,1)</td>
<td>(-L) without review, (-2L) with review</td>
<td>(-L) without review, (-2L) with review</td>
</tr>
<tr>
<td>(\theta = 0 ) \ (-1,0;0,0)</td>
<td>(0) without review, (0) with review</td>
<td>(0) without review, (0) with review</td>
</tr>
<tr>
<td>(\theta = -1 ) \ (0,0;0,0)</td>
<td>(0) without review, (0) with review</td>
<td>(0) without review, (0) with review</td>
</tr>
</tbody>
</table>

Thus the Supreme Court’s utility without review is \(-L\). The Supreme Court’s utility if it reviews is \(-c - 2L \times Pr(\theta = -1 + 1; -1,0;0,0\mid A,M)\). Therefore the Supreme Court will
After \((A, B)\) the Supreme Court’s beliefs are \(Pr(\theta = 0) = 1\). Its utility without review is \(-2L\). Its utility upon review is \(-c\). The Supreme Court reviews whenever \(c < 2L\).

**Proof.** Ideological Heterogeneity.

1. **The Lower Courts** \(LC_I\) behaves as the lower courts did in Ideological Homogeneity.

   After \(LC_{II}\) sees \((0,1)\) his expected utility from each possible decision is:
   
   \[
   EU_{LC_{II}}[B] = -1 \ast Pr(\theta = 0|0, 1) = -\alpha
   \]
   
   \[
   EU_{LC_{II}}[M] = -L \ast Pr(\theta = 1|0, 1) = -L(1 - \alpha)
   \]
   
   \[
   EU_{LC_{II}}[A] = -1
   \]

   So \(LC_{II}\) chooses \(M\) so long as \(L < \frac{\alpha}{1-\alpha}\), which I assume henceforth.\(^8\) Thus after seeing \((0,1)\), \(LC_{II}\) chooses \(M\) instead of \(B\). After seeing \((-1,0)\), he chooses \(A\).

2. **The Supreme Court** After observing \((A,A)\); \((B,A)\); or \((M,A)\) the Supreme Court’s beliefs and strategies are as in Ideological Homogeneity. After observing \((A,M)\); \((M,M)\); or \((B,M)\) the Supreme Court’s beliefs are different.\(^9\)

   **After \((A, M)\):** The Supreme Court knows \(\theta_A = -1\) but does not know if \(\theta_B = 0\) or \(= 1\).

   It can review \(LC_{II}\) to learn this: if it observes \((0,0)\) it updates the probability that \(\theta_B = 0\); if it observes either \((-1,1)\) or \((0,1)\) it concludes with certainty that \(\theta_B = 1\). The Supreme

---

\(^8\)It would also be acceptable to increase the loss from \(B\) to a loss still greater than 1. I choose to manipulate \(L\) instead for algebraic simplicity.

\(^9\)After observing \((A,B)\); \((M,B)\); or \((B,B)\) beliefs are off-path; I assume that the Supreme Court believes \(LC_{II}\) received messages \((0,1)\).
Court loses $L$ from each decision of $M$ with $|\theta| = 1$ and $L$ from each decision of $A$ or $B$ with $\theta = 0$. Without reviewing either case, the Supreme Court’s expected utility is:

$$-L \cdot Pr(\theta = 0) - L \cdot Pr(\theta = -1) - 2L \cdot Pr(\theta = 1) = -L.$$ 

Thus its utility is $-L$ (because one court must be right and the other must be wrong).

Upon review, it will learn $LC_{II}$ observed either $(0,0)$, $(-1,1)$, or $(0,1)$. After either $(-1,1)$ or $(0,1)$ it will know $\theta = 0$ with certainty and will be able to achieve utility 0 by setting doctrine $M$. If it learns $(0,0)$, it will know $\theta_A = -1$ and will believe more strongly that $\theta_B = 0$, but will still not know this with certainty and will incur some loss from choosing $A$ despite the possibility that $\theta = 0$.

<table>
<thead>
<tr>
<th>State of the world</th>
<th>Messages</th>
<th>Utility change</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta = -1$</td>
<td>$(-1,0,0,0)$</td>
<td>-L without review, 0 with review</td>
</tr>
<tr>
<td>$\theta = -1$</td>
<td>$(-1,0,0,1)$</td>
<td>Cannot happen</td>
</tr>
<tr>
<td>$\theta = -1$</td>
<td>$(-1,0,-1,1)$</td>
<td>Cannot happen</td>
</tr>
<tr>
<td>$\theta = -1 + 1$</td>
<td>$(-1,0,0,0)$</td>
<td>-L without review, -2L with review</td>
</tr>
<tr>
<td>$\theta = -1 + 1$</td>
<td>$(-1,0,0,1)$</td>
<td>-L without review, 0 with review</td>
</tr>
<tr>
<td>$\theta = -1 + 1$</td>
<td>$(-1,0,-1,1)$</td>
<td>-L without review, 0 with review</td>
</tr>
</tbody>
</table>

Therefore its utility from review is $-c - 2LP r(\theta = -1 + 1; -1,0,0,0|B,M)$. It will review so long as

$$-c - 2LP r(\theta = -1 + 1; -1,0,0,0|B,M) > -L$$ 
$$L - 2LP r(\theta = -1 + 1; -1,0,0,0|B,M) > c$$
$$c < L - 2LP r(\theta = -1 + 1; -1,0,0,0|B,M)$$
$$c < L[1 - 2Pr(\theta = -1 + 1; -1,0,0,0)] Pr(B,M)$$
$$c < L[1 - 2\frac{\alpha(1-q)^2}{(1-\alpha)(1-q)^2+\alpha(1-q)^2+\alpha q(1-q)+\alpha q^2}]$$

After $(B, M)$: The Supreme Court knows $\theta_B = 1$ but does not know whether $\theta_A = 0$ or $= -1$. It can review $LC_{II}$ to learn this: if it observes either $(0,0)$ or $(0,1)$ its posterior belief that $\theta_A = 0$ grows stronger; if it observes $(-1,1)$ it concludes with certainty that $\theta_A = -1$. 

38
Its utility gain from reviewing is as follows:

<table>
<thead>
<tr>
<th>State of the world</th>
<th>Messages</th>
<th>Utility change</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta = 1 )</td>
<td>((0,1;0,0))</td>
<td>-L without review, 0 with review</td>
</tr>
<tr>
<td>( \theta = 1 )</td>
<td>((0,1;0,1))</td>
<td>-L without review, 0 with review</td>
</tr>
<tr>
<td>( \theta = 1 )</td>
<td>((0,1;-1,1))</td>
<td>Cannot happen</td>
</tr>
<tr>
<td>( \theta = -1 + 1 )</td>
<td>((0,1;0,0))</td>
<td>-L without review, -2L with review</td>
</tr>
<tr>
<td>( \theta = -1 + 1 )</td>
<td>((0,1;0,1))</td>
<td>-L without review, -2L with review</td>
</tr>
<tr>
<td>( \theta = -1 + 1 )</td>
<td>((0,1;-1,1))</td>
<td>-L without review, 0 with review</td>
</tr>
</tbody>
</table>

Therefore it will review if:

\[-L < -c - 2L \times Pr(\theta = -1 + 1; 0, 1; 0, |B, M)\]
\[-L < -c - 2L \alpha^2(1-q)^2q\]
\[c < L \times 2L \alpha(1-\alpha)q(1-q)+q^2+\alpha^2[q(1-q)^3+q^2(1-q)^2+q^3(1-q)]\]
\[c < L \times 2L \alpha(1-\alpha)(1-q)(1-q+q^2)\]

After (M, M): The Supreme Court puts positive probability on all values of \( \theta \). The utility change from reviewing is as follows:
\[ \begin{array}{|c|c|c|} \hline \text{State of the world} & \text{Messages} & \text{Utility change} \\ \hline \theta = -1 & (0,0,0) & \text{-2L without review, -2L with review} \\ \theta = -1 & (0,0,1) & \text{Cannot happen} \\ \theta = -1 & (0,0;-1,1) & \text{Cannot happen} \\ \theta = -1 & (-1,0,0) & \text{Cannot happen} \\ \theta = -1 & (0,1,0) & \text{Cannot happen} \\ \theta = -1 & (-1,1;0,0) & \text{Cannot happen} \\ \theta = 1 & (0,0,0) & \text{-2L without review, 0 with review} \\ \theta = 1 & (0,0,1) & \text{-2L without review, 0 with review} \\ \theta = 1 & (0,0;-1,1) & \text{Cannot happen} \\ \theta = 1 & (-1,0,0) & \text{Cannot happen} \\ \theta = 1 & (-1,1;0,0) & \text{Cannot happen} \\ \theta = 1 & (-1,1;0,1) & \text{Cannot happen} \\ \theta = 1 & (-1,1;-1,1) & \text{Cannot happen} \\ \theta = -1 + 1 & (0,0,0) & \text{0 without review, 0 with review} \\ \theta = -1 + 1 & (0,0,1) & \text{0 without review, -2L with review} \\ \theta = -1 + 1 & (0,0;-1,1) & \text{0 without review, 0 with review} \\ \theta = -1 + 1 & (-1,0,0) & \text{0 without review, 0 with review} \\ \theta = -1 + 1 & (-1,1;0,1) & \text{0 without review, -2L with review} \\ \theta = 0 + 0 & (0,0,0) & \text{0 without review, 0 with review} \\ \theta = 0 + 0 & (0,0,1) & \text{Cannot happen} \\ \theta = 0 + 0 & (0,0;-1,1) & \text{Cannot happen} \\ \theta = 0 + 0 & (-1,1,0,0) & \text{Cannot happen} \\ \theta = 0 + 0 & (-1,1,0,1) & \text{Cannot happen} \\ \theta = 0 + 0 & (-1,1;-1,1) & \text{Cannot happen} \\ \hline \end{array} \]

Therefore, the Supreme Court’s expected utility without review is
\[
-2L \cdot Pr(\theta = 1; 0, 0; 0, 1 | M, M) = -2L \frac{\alpha (1-\alpha)q(1-q)}{Pr(M,M)}
\]
\[
= -2L \frac{\alpha (1-\alpha)q(1-q)}{(1-\alpha)^2 + \alpha (1-\alpha)(1-q)^2 + \alpha (1-\alpha)(1-q) + \alpha^2 [(1-q)^2 + (1-q)^2 q^2 + q^2 (1-q) + q^4]}
\]

The Supreme Court’s expected utility with review is
\[
-c - 2L \frac{\alpha^2 q (1-q)^3 - \alpha^2 q^3 (1-q)}{(1-\alpha)^2 + \alpha (1-\alpha)(1-q)^2 + \alpha (1-\alpha)(1-q) + \alpha^2 [(1-q)^2 + (1-q)^2 q^2 + q^2 (1-q) + q^4]}
\]
The Supreme Court will review if:

\[-c < \frac{2L\alpha q(1-q)^3 - 2L\alpha^2 q^3(1-q)}{(1-\alpha)^2 + \alpha(1-\alpha)(1-q)^2 + \alpha(1-\alpha)(1-q) + \alpha^2[(1-q)^3 + (1-q)^4 + 2(1-q)^2q^2 + q^4(1-q) + q^4]}\]

Otherwise it will not review. ■
Chapter 3

Review

The model generates a number of empirical predictions. To begin, I focus on analyzing which cases the Supreme Court chooses to review. In particular, the theory suggests moderate decisions—where each party prevails on some counts—provide much more information than decisions where one party prevails on all counts. If the Supreme Court is attempting to learn, and to develop doctrine, it will be more likely to review these cases.

Because the Supreme Court takes so few cases each year, the choice of which cases to review is one of the most important in understanding the Supreme Court’s role. Therefore, this choice—the decision of whether to grant *certiorari* and review a case—has been studied extensively. The literature on *certiorari* is too vast to be discussed comprehensively here. Studies have focused primarily on two factors: first, the ideology of lower court judges who made a decision, and second, external cues that suggest a case is difficult, incorrectly decided, particularly important, or otherwise worthy of review. With respect to ideology, scholars have argued that the Supreme Court is more likely to review ideologically distant lower court judges ([Cameron, Segal and Songer 2000](#), [Lindquist, Haire and Songer 2007](#), [Black and Owens 2009](#), [Hall 2009](#)) and that the Supreme Court is more likely to review judges who are not too...
ideologically distant nor too ideologically proximate. Participation of outside actors, such as involvement of the solicitor general (see e.g. Bailey, Kamoie and Malztman 2005, Black and Owens 2012) or of interest groups participating as amici curiae (Caldeira and Wright 1988) also increase the likelihood of review. (Both of these, importantly, occur very infrequently and lead to review at extremely high rates.)

As previously discussed, indicators of legal ambiguity or difficult, such as conflict between the circuits, has a dramatic effect on the probability of certiorari, but because conflict is so much more common, it leads to Supreme Court review with much less frequency (Estreicher and Sexton 1984). The same is true of cues that a decision is worthy of reversal (Brenner and Krol 1989, Krol and Brenner 1990), such as a dissent or a reversal of the district court’s decision (Tanenhaus et al. 1963, Perry 1991, Caldeira, Wright and Zorn 1999).

Crucially, however, one major component is missing from these analyses: the direction of the lower court’s decision. Those studies that do consider a disposition’s effect on cert ask only whether the Supreme Court is more likely to review liberal or conservative decisions made in the lower courts. As will be shown below, this dichotomy obfuscates the importance of decisional ambiguity in predicting certiorari. If the Supreme Court is policing lower courts to ensure compliance, as most previous literature has argued, it will be less likely to review moderate decisions, and more likely to review extreme decisions. Using a dataset of about 6,000 Courts of Appeals decisions, in this chapter I show the Supreme Court is indeed more likely to review moderate than extreme decisions. The results suggest that the Court is not primarily concerned with ensuring that lower courts follow established doctrine—that is, with deciding easy cases as the Supreme Court would like—but rather with the establishment and extension of new doctrine—that is, with the development of new rules for disposing of future cases.
3.1 Theory and Hypotheses

Which lower court decision will the Supreme Court choose to review (A, M, or B, representing what one could think of as liberal, moderate, and conservative decisions)? The model predicts the Supreme Court will review moderate decisions above extreme ones. Specifically, the model predicts the following:

**Moderate Cases Hypothesis:** The Supreme Court will be more likely to review decisions where each party wins on some counts, than decisions where one party wins on all counts.

This prediction brings to the forefront the conflict between an incentive to learn and an incentive to discipline. Whereas a purely ideological model focused on discipline would predict a conservative Supreme Court would target liberal lower courts’ liberal decisions for review and reversal, the learning model predicts even a conservative Supreme Court will target moderate decisions for review. To review the intuition behind this prediction, recall that when one lower court has made a decision of M and the other has not, if the Supreme Court reviews, it will review the decision of M. Decisions of A will only be reviewed if the other lower court chooses B, and even then review is not certain. Furthermore, recall from Proposition 2 that LC11’s decisions of M are most likely to be reviewed (compared to LC1’s decisions and LC11’s other decisions). This is because all other decisions are reviewed only conditional on the other courts’ ruling. Assuming review is not too costly, it follows that the Supreme Court is more likely to review moderate decisions than extreme decisions, and that in particular the Supreme Court is most likely to review mixed decisions made by biased lower courts.
3.2 Data

The data for this analysis are derived from 6,971 Courts of Appeals cases decided between 1970 and 1986. The data are described in detail in the appendix. Briefly, the data are a choice-based sample [Xie and Manski 1989, King and Zeng 2001] of Courts of Appeals decisions, including all cases decided by three-judge panels in the Courts of Appeals that the Supreme Court reviewed and a stratified random sample of cases that were not reviewed. I took the key variables—the judges on the panel, the decision the panel reached, and whether the Supreme Court reviewed the decision—from the Songer Phase I and Phase II Courts of Appeals databases [Songer 1999, 2008], merged by Clark and Carrubba (2012). The dependent variable, the *certiorari* decision—that is, whether the Supreme Court chose to review or not to review the panel’s ruling—is straightforward. Coding the independent variables is somewhat more complex.

In the model, there are three possibilities for a lower court’s decision—one party wins in the lower court, the other party wins, or each prevails on some counts. Likewise, in the Songer Databases, the decision of the lower court is coded based on which party wins. In general, a decision is coded as *Liberal* if the liberal party wins (such as the defendant in a criminal trial, or a labor union pursuing an action against management) and is coded as *Conservative* if the conservative party wins (such as the state in a criminal trial, or an employer). When each party prevails on some claims, the decision is coded as *Mixed*.

---

1. I dropped all records the Supreme Court consolidated for review because the model assumes the Supreme Court can take only one case. Additionally, the data does not include cases heard by full circuits sitting *en banc*—only cases heard by three-judge panels.
2. For some types of cases in which the liberal party is hard to identify, such as conflicts between rival unions or commercial disputes in which no party is clearly an underdog, the outcomes are coded as unspecifiable. There are 885 such cases in the data, including 13 reviewed cases for a weighted mean of .2%. Because these unspecifiable decisions are almost never reviewed, I remove them from the sample.
3. For a helpful discussion of when mixed decisions arise in the Courts of Appeals, see Lindquist, Haire and Songer (2007).
panel will occasionally rule on two issues in one case, finding for the liberal party on one issue and for the conservative party on the other. Other times, there is only one issue in the case, and the panel will issue a liberal, mixed, or conservative ruling on that single question. I consider both possibilities for identifying mixed decisions: first coding as mixed only those cases that have a mixed outcome on a single dimension, and following Landes and Posner (2009) and coding as mixed any decision in which the panel voted liberally on one issue and conservatively on another. The appendix shows the distribution of decisions across panel types.

To translate the model to data, I define a biased lower court as one whose ideology is different from the Supreme Court’s. When the Supreme Court is conservative, for example, a liberal lower court is a biased lower court. Measuring judicial ideology is a notoriously thorny problem. This is particularly true when trying to place judges from different courts on the same scale. To get around this, I use an indirect approach that several studies have also employed (see for example Hall (2009), Cameron, Segal and Songer (2000), and Kastellec (2011)). For the Supreme Court, I assume a fixed conservative ideology. Although the Supreme Court did become more conservative between 1970 and 1986, its ideology remained conservative (see for example Bailey (2007)).

I measure lower court judges’ ideology by the party of the president that nominated them.4 Because the rate of unanimous decisions is so high on the Courts of Appeals, a large literature on panel effects in the lower courts suggests that measures such as ideology of the opinion’s author or the median judge may be misleading (e.g. Sunstein et al. (2006)), so it is necessary to account for complete panel composition. Therefore, I separately consider each

4Broadly, this measure is accurate, as presidents tend to appoint like-minded judges to the bench, but it is obviously imperfect. The measure is likely to perform worst in the South, where some Democratic-appointed judges in this time period are more conservative than their Republican counterparts in the North. Below, I show that results continue to hold considering only non-southern circuits (that is, excluding the 5th and 11th circuits).
panel type—all Democratic nominees sitting together (abbreviated DDD), two Democrats sitting with one Republican (DDR), two Republicans sitting with one Democrat (RRD), or three Republican nominees sitting together (RRR). The important assumption is that the median Supreme Court justice—Justice Byron White, for most of this time period (Martin and Quinn 2002)—was more conservative than Democratic-appointed Courts of Appeals judges. Under this assumption, liberal panels—DDD panels—are considered “biased” lower courts, since the Supreme Court is assumed to be conservative. Conservative panels—RRR panels—are “unbiased” lower courts.

In the analysis, I also include an indicator of whether one lower court judge Dissented from the panel’s ruling, as dissent is known to predict Supreme Court review (Tanenhaus et al. 1963, Perry 1991, Caldeira, Wright and Zorn 1999). Additionally, I control for the issue area of the case, since the Supreme Court may be more likely to hear some types of cases than others. A case is either coded as a Criminal case, or as concerning Economic activity and regulation, or as neither of these categories. Residual cases include civil rights, First Amendment, due process, privacy, labor relations, and other cases. I also control for the Number of amicus briefs filed at the Courts of Appeals, to proxy for the salience or importance of the case. Additionally, since the Supreme Court may be more likely to review some circuits than others for reasons not accounted for by the model, I include (but do not present) fixed effects for Circuit, and, since the rate of review may have changed over time, I control for the year of the decision (Time). Finally, because the data includes all cases that were reviewed, and a random sample of cases that were not reviewed, all analyses weight to correct for choice-based sampling (Xie and Manski 1989, King and Zeng 2001).
3.3 Panel Composition and Decisions

Beginning with the raw data, decisions where each party wins on some counts are more likely to be reviewed than those where one party wins on all counts. 3.4% of liberal decisions are reviewed and only 1.5% of conservative decisions are reviewed; in contrast, 5.7% of mixed decisions are reviewed. In fact, mixed decisions make up 22% of cases the Supreme Court reviews, even though they make up only 12% of cases decided in the Courts of Appeals.

Table 3.1 shows the distribution of cases in the full data, and among those cases that were reviewed by the Supreme Court. Within the lower courts, Conservative and Liberal decisions are made at approximately equal rates; mixed decisions are much less frequent. (Recall, though, that the universe of cases (All) is not a random sample of cases decided in the Courts of Appeals.) Most of the cases the Supreme Court chooses to review were liberal decisions made in the lower courts, which is consistent with the notion that the Burger Court was largely conservative and with a perspective of the Supreme Court as reviewing in order to reverse lower courts’ decisions. As will be shown, however, this does not fully explain Supreme Court review behavior.

<table>
<thead>
<tr>
<th></th>
<th>Unspecifiable</th>
<th>Conservative</th>
<th>Mixed</th>
<th>Liberal</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>8.6%</td>
<td>43.7%</td>
<td>12.0%</td>
<td>35.7%</td>
</tr>
<tr>
<td>Reviewed</td>
<td>3.7%</td>
<td>29.2%</td>
<td>22.4%</td>
<td>44.7%</td>
</tr>
</tbody>
</table>

Table 3.1: The distribution of decision direction among all cases (top row) and among those that are reviewed by the Supreme Court (bottom row).

Table 3.2 shows comparable figures for panel composition. It is noteworthy, although unsurprising, that mixed panels (with some Democratic appointees and some Republican appointees) are much more common than unified panels.\(^5\) It is also noteworthy that the

\(^5\)See Kastellec (2011b) for a discussion of distribution across panel compositions over time.
Table 3.2: The distribution of panel composition among all cases (top row) and among those that are reviewed by the Supreme Court (bottom row).

<table>
<thead>
<tr>
<th></th>
<th>DDD</th>
<th>DDR</th>
<th>RRD</th>
<th>RRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>13%</td>
<td>41%</td>
<td>34%</td>
<td>12%</td>
</tr>
<tr>
<td>Reviewed</td>
<td>14%</td>
<td>39%</td>
<td>33%</td>
<td>15%</td>
</tr>
</tbody>
</table>

distribution of panels is similar in the cases that are reviewed and the universe of cases. In other words, all panels are reviewed at approximately equal rates. This suggests the Supreme Court is not more likely to review panels it sees as ideologically distant. But, Table 3.1 shows that different decisions are reviewed at markedly different rates. These differences yield the results discussed below—which cases does the Supreme Court choose to review?

### 3.4 Results

To analyze Supreme Court review decisions, I run a set of logistic regressions of *certiorari* on panel composition and decision direction. Table 3.3 presents the coefficients from these regressions. Model 1 includes only decision direction, and replicates the statistics described above—compared to the omitted category (conservative decisions), mixed decisions are much more likely to be reviewed. Liberal decisions are also more likely to be reviewed than conservative decisions, a point I return to below.

Turning to Model 2, which interacts decision direction with panel composition, it is evident that, when they make conservative decisions, DDD panels are *less* likely to be reviewed than their RRR counterparts. This is a standard finding in the literature on *certiorari*, and a focus of Cameron, Segal and Songer (2000): the Supreme Court can be certain it will agree when a court more liberal than it makes a conservative decision. Liberal decisions are still more likely to be reviewed than conservative decisions, the omitted category. The normalization of the Supreme Court as an unbiased supervisor prevents the model from speaking to
such a prediction, but given that the Supreme Court is conservative during this time period, this effect makes sense—though it is notable that the effect is seen in the *decisions*, not the panel ideology.

The remaining regressions include a number of categorical independent variables, and so the coefficients must be interpreted with care. In Model 5, for example, the omitted category is unanimous conservative decisions by unified Republican panels in the First Circuit, with no amicus briefs submitted at the Courts of Appeals stage, in cases that are neither criminal nor economic. Each coefficient shows the change in probability of review relative to this baseline. I translate the coefficients into predicted probabilities below, but some effects are clear from the coefficients themselves.

Model 3 restricts the Mixed coding to those cases which were found to be Mixed on a singular dimension. The effect is still present—Mixed decisions are still most likely to be reviewed—and we thus see that the effect is not attributable to an increased complexity of Mixed decisions. Even when cases are relatively simple, if the decision on the merits is moderate, the case is more likely to be reviewed. Model 4 excludes the Southern 5th and 11th circuits, where party is likely to be a weaker proxy for ideology in this time period. Results get stronger still—the effect of Mixed decisions is positive *and* the interaction between Mixed and comparatively biased DDD panels is statistically significant.

Turning to Model 5 (which includes all circuits and considers both issues in a case when two exist) we see that the effect of Mixed decisions remains positive and statistically significant and that including relevant control variables makes all effects of panel ideology statistically indistinguishable from 0. Decision direction is a much more powerful predictor of review than is panel ideology. Most importantly, the total effect of mixed decisions by liberal panels—1.4—is significantly larger than the total effect of liberal decisions by liberal panels—.43.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Two Issues</td>
<td>Two Issues</td>
<td>Primary Issue</td>
<td>Non-south only</td>
<td>Two Issues</td>
</tr>
<tr>
<td>Mixed</td>
<td>1.36*</td>
<td>0.93*</td>
<td>0.80*</td>
<td>0.79*</td>
<td>0.77*</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.27)</td>
<td>(0.37)</td>
<td>(0.33)</td>
<td>(0.34)</td>
</tr>
<tr>
<td>Liberal</td>
<td>0.82*</td>
<td>0.49*</td>
<td>0.51</td>
<td>0.53*</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.20)</td>
<td>(0.26)</td>
<td>(0.27)</td>
<td>(0.27)</td>
</tr>
<tr>
<td>DDD</td>
<td>-0.42*</td>
<td>0.01</td>
<td>-0.10</td>
<td>-0.08</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.28)</td>
<td>(0.30)</td>
<td>(0.29)</td>
<td></td>
</tr>
<tr>
<td>DDR</td>
<td>-0.40*</td>
<td>-0.03</td>
<td>-0.06</td>
<td>-0.08</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.24)</td>
<td>(0.25)</td>
<td>(0.25)</td>
<td></td>
</tr>
<tr>
<td>RRD</td>
<td>-0.50*</td>
<td>-0.24</td>
<td>-0.27</td>
<td>-0.32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.24)</td>
<td>(0.24)</td>
<td>(0.25)</td>
<td></td>
</tr>
<tr>
<td>DDD, Mixed</td>
<td>0.80*</td>
<td>0.55</td>
<td>0.81*</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>(0.37)</td>
<td>(0.45)</td>
<td>(0.43)</td>
<td>(0.42)</td>
<td></td>
</tr>
<tr>
<td>DDR, Mixed</td>
<td>0.39</td>
<td>0.21</td>
<td>0.32</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.38)</td>
<td>(0.35)</td>
<td>(0.35)</td>
<td></td>
</tr>
<tr>
<td>RRD, Mixed</td>
<td>0.53</td>
<td>0.30</td>
<td>0.39</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.41)</td>
<td>(0.36)</td>
<td>(0.36)</td>
<td></td>
</tr>
<tr>
<td>DDD, Liberal</td>
<td>0.30</td>
<td>-0.07</td>
<td>0.11</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.33)</td>
<td>(0.36)</td>
<td>(0.34)</td>
<td></td>
</tr>
<tr>
<td>DDR, Liberal</td>
<td>0.37</td>
<td>0.06</td>
<td>0.16</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.26)</td>
<td>(0.28)</td>
<td>(0.27)</td>
<td></td>
</tr>
<tr>
<td>RRD, Liberal</td>
<td>0.47*</td>
<td>0.17</td>
<td>0.22</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(0.28)</td>
<td>(0.29)</td>
<td>(0.29)</td>
<td></td>
</tr>
<tr>
<td>Dissent</td>
<td>-0.56*</td>
<td>-0.54*</td>
<td>-0.55*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>-0.24*</td>
<td>-0.31*</td>
<td>-0.25*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number Amicus Briefs</td>
<td>0.34*</td>
<td>0.32*</td>
<td>0.33*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.14*</td>
<td>-3.74*</td>
<td>-3.70*</td>
<td>-3.79*</td>
<td>-3.68*</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.13)</td>
<td>(0.36)</td>
<td>(0.37)</td>
<td>(0.37)</td>
</tr>
<tr>
<td>Circuit Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>6,289</td>
<td>6,289</td>
<td>6,289</td>
<td>5,689</td>
<td>6,289</td>
</tr>
</tbody>
</table>

Table 3.3: The probability of review, by panel composition and decision direction. Standard errors in parentheses. * indicates $p < .05$. All analyses performed using the logit.survey package in Zelig (Carnes 2007), with weights to correct for choice-based sampling.

Figure 3.1 translates these effects into predicted probabilities of review, by decision and panel type. The figure makes it clear that decision type is a far more influential variable than
panel composition. Mixed decisions are by far the most likely cases to be reviewed, regardless of the court that made them. Even though decision direction is more powerful a predictor than is court ideology, among those who make mixed decisions biased judges are still more likely to be reviewed. As the model predicts, the mixed decisions of ideologically distant panels (DDD) are most likely to be reviewed. These liberal panels’ liberal decisions have a 3% chance of review; liberal panels’ mixed decisions have an 8% probability of review. This is a sizable increase; the difference between two rates is distinguishable from 0 at \( p < .05 \). This is also particularly noteworthy given that review is so rare—the overall probability of review is a mere 2.5%.

3.5 Discussion and Conclusion

The Supreme Court is not most likely to review extreme decisions. Instead, it focuses its attention on reviewing cases that are ambiguous and suggestive of law that needs development, and in particular on those such cases that provide information enabling immediate legal progress.

In contrast with many existing theories of the judicial hierarchy, the results show the conservative Burger Court was not most likely to review liberal courts’ liberal decisions, as predicted by a purely ideological framework. Instead, it was most likely to review the decisions that were likely to have been difficult and will therefore offer learning opportunities. Previous literature on the judicial hierarchy would argue that a conservative Supreme Court wishes to minimize liberal decisions in the lower courts, and therefore targets those for review. Empirically, however, the relatively conservative Supreme Court targeted mixed rather than purely liberal decisions. The results also highlight the importance of not dichotomizing lower court dispositions. Those cases in which only one party wins on all counts are likely to be more straightforward applications of law, or likely to be simpler for outside actors to
interpret; thus, as a methodological point, the results suggest coding decisions trichotomously is important.

This result is broadly consistent with Carrubba and Clark (2012), who find that panels are most likely to be reviewed when they are not too allied and not too distant from the U.S. Supreme Court. The results suggest that the Supreme Court is not primarily concerned with lower courts following established liberal doctrine, but with preventing the establishment and extension of new liberal doctrine. The Supreme Court wishes to minimize liberal lawmaking in the lower courts, as opposed to decisions.
Appendix: Data Sources

To construct the data for the review analysis, I began with the data from Carrubba and Clark (2012). That dataset was composed as follows. For every year from 1961 to 1986 the raw dataset included 30 cases from each circuit, randomly sampled. This was taken from Phase I of the Courts of Appeals Database (Songer 1999). Any cases the Supreme Court reviewed that were not captured in this random sample were then collected and added in Phase II of the database (Songer 2008). To connect the Courts of Appeals cases to Supreme Court cases, Carrubba and Clark (2012) merged the data with the Original Spaeth Supreme Court Database (Spaeth 2011). The resulting Carrubba and Clark (2012) dataset included all Courts of Appeals cases from 1961 to 1986 that the Supreme Court reviewed, plus a stratified random sample of cases that were not reviewed.

I then did the following. First, I selected only cases between 1970 and 1986, when the Supreme Court had a stable conservative stance (see Bailey (2007)). I also dropped all records associated with a consolidated decision by the Supreme Court, because the model assumes the Supreme Court can take only one case. Then, I corrected a number of mistakenly identified Courts of Appeals judges and added information on judges sitting by designation, using data from Kastellec (2011b). The finalized, clean data includes 6,971 cases.

Because the cases are a stratified random sample, all analyses weight observations to simulate a random sample from the Courts of Appeals. Weights are constructed based on population estimates in Songer (1999). The sample includes more reviewed cases than would be sampled at random, and more cases from small circuits than would be sampled at random, so these have weights less than 1. The sample includes fewer unreviewed cases than would be sampled at random, so these are given weights greater than 1.

For data on the party of lower court judges’ nominating president, I use Gryski and Zuk
(2008) and Gryski, Zuk and Goldman (2008). A judge is identified as D if he was nominated by a Democratic president, and R if he was nominated by a Republican president. See text for description of coding of case outcomes. Table 3.4 presents the distribution of panel composition by case outcome.

<table>
<thead>
<tr>
<th>Unspecifiable</th>
<th>Conservative</th>
<th>Mixed</th>
<th>Liberal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDD</td>
<td>.01</td>
<td>.05</td>
<td>.02</td>
<td>.05</td>
</tr>
<tr>
<td>DDR</td>
<td>.03</td>
<td>.17</td>
<td>.05</td>
<td>.15</td>
</tr>
<tr>
<td>RRD</td>
<td>.03</td>
<td>.16</td>
<td>.04</td>
<td>.12</td>
</tr>
<tr>
<td>RRR</td>
<td>.01</td>
<td>.06</td>
<td>.01</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>587</strong></td>
<td><strong>3054</strong></td>
<td><strong>834</strong></td>
<td><strong>2496</strong></td>
</tr>
</tbody>
</table>

Table 3.4: Distribution of panel types and decision direction. Recall that sampling is non-random, so analyses are performed with weights.
Chapter 4

Conflict

When it decided *NFIB v. Sebelius* in 2012, the Supreme Court entered one of the nation’s most controversial policy debates of the time: the national healthcare proposal widely known as Obamacare. Four Courts of Appeals—the Fourth, Sixth, Eleventh, and D.C. circuits—had previously heard cases contesting the constitutionality of the healthcare act. In each of the four cases, litigants filed petitions for *certiorari*, asking the Supreme Court to hear the case. The Supreme Court could have consolidated the cases and granted *cert* to all of them, but in this instance it chose not to—it granted *cert* to the Eleventh Circuit and denied to the Fourth, Sixth, and D.C. Circuits. Upon review, the Supreme Court reversed the Eleventh Circuit’s decision, and issued a decision that was in many respects consistent with those of the Fourth, Sixth, and D.C. Circuits. Did review of the Eleventh Circuit’s decision, as opposed to another decision, impact the Court’s opinion? Had the Supreme Court heard a different circuit’s case, would its decision have been different?

---

1The petition for certiorari from the Eleventh Circuit was not the first to arrive to the Supreme Court. The petition for certiorari from the Sixth Circuit’s decision, 651 F.3d 529, was filed first, on July 26, 2011 (docket number 11-117). The petitions from the Eleventh were filed on September 28, 2011 (docket numbers 11-393, 11-398, and 11-400); the Fourth Circuit petition was filed on October 7, 2011 (docket number 11-438), and the DC Circuit petition was filed on November 30, 2011 (docket number 11-679). The other petitions were only rejected on June 29, 2012, after *NFIB v. Sebelius* had been decided.
Although the healthcare case was extremely high-profile, it is not uniquely so. In recent decades, the U.S. Supreme Court has been at the forefront of controversial policy-making: it has been a critical actor in the desegregation of public schools, the development of abortion policy, the evolution of rights for same-sex couples, and the construction of national healthcare policy. But unlike the President and Congress, Supreme Court policymaking is unique, in that it is not abstract. The Supreme Court can only step in when a case or controversy is presented to it; but along with this the Supreme Court also has the liberty to choose which case or controversy to review. As a result, there is a close relationship between the cases the Court chooses to review, and the doctrines it decides to articulate.

Scholars have proposed many ways by which review and doctrine might be related. The Supreme Court may target doctrines it disagrees with (Bonneau et al. 2007), or may focus only on unusual cases that further doctrinal development (Kastellec and Lax 2008), or may choose to review cases to adopt opinion content (Clark and Carrubba 2012, Corley, Collins and Calvin 2011). The relationship between review and reversal may also be influenced by the particular arguments brought forth in particular cases. If this is the case, the Supreme Court’s review decisions will be contextual in a way not previously proposed—the Supreme Court will use other decisions to inform its choice of which case to review.

Even very similar cases present different arguments and information. Bashman (2013), an appellate lawyer, describes arguing a case of first impression before the Third Circuit. A case similar to his own reached the Third Circuit of the Courts of Appeals shortly before his would have. But the lawyers in the related case failed to present the arguments Bashman found relevant, so he feared the circuit is stuck with imperfect law because those lawyers failed to present information which he would have presented. Therefore, the Supreme Court’s choice of which decision to review—even when all cases address the same question—is of critical importance for the broader development of legal doctrine.
In this chapter, I evaluate the relationship between review and doctrine based on the model’s predictions. I consider whether the Supreme Court is more likely to support the doctrine set forth in the cases it chooses to review or in cases it chooses not to review; in so doing, I show how the Supreme Court’s doctrinal choices are influenced by lower courts’ decisions.

4.1 Measuring Doctrine

The theory of learning from lower courts offers a prediction that is unusual in models of the judicial hierarchy: rather than speaking only to the Supreme Court’s dispositional treatment of the lower court, the model also makes predictions about what doctrine the Supreme Court will choose. In many respects, this doctrinal choice is the more important quantity, for while litigants may care about the disposition of one particular case, others typically care about the doctrine articulated in its resolution. But while doctrine might be considered the more meaningful of the Supreme Court’s actions, it is also the most difficult to empirically quantify. Assessing the Supreme Court’s stated doctrine and its relationship to lower court doctrine is a complicated empirical challenge. With a few notable exceptions, research on the relationship between the Supreme Court and the Courts of Appeals has focused on the Supreme Court’s dispositional treatment of the courts below.

Quantifying Supreme Court doctrine is an inherently complex task. There are currently two popular solutions to this difficulty: qualitative doctrinal tracing, and automated text analysis. The latter, which has become increasingly popular in recent years, uses citations to quantify the location of a Supreme Court opinion. These strategies have shown how

\(^{2}\)Although empirical studies of doctrine and its relationship to the hierarchy are rare, this is not the first such study—see for example Clark and Carrubba (2012) and Corley, Collins and Calvin (2011).
the Supreme Court’s citation choices are influenced by lower courts and previous Supreme Court decisions, which provides substantial insight into doctrinal origins. It is still somewhat unclear whether citation patterns reflect doctrinal choices, though. An alternative strategy is to use text analysis of the opinions themselves, as in Corley, Collins and Calvin (2011). However, this strategy tends to equate the facts of the case with its coincident doctrine, so that it is difficult to tell what the Supreme Court is mimicking when it employs similar language. Furthermore, within a legal issue, words are so similar to one another that natural language processing cannot identify correlations between words and doctrine (Beim 2012).

Instead, I use an alternative approach. I rely on conflicts between the circuits to test the hypothesis using the Supreme Court’s own stated opinions. When the Supreme Court resolves a conflict between two or more lower courts, the majority opinion describes the conflict and explicitly states which side (if either) it is supporting. Thus, conflicts allow a researcher to examine the relationship between the Supreme Court’s doctrine and lower court doctrine without having to develop a metric.

This has similarities to the citation-based methods employed in Clark and Lauderdale (2010), Hansford and Spriggs (2008), Hansford, Spriggs and Stenger (2010) and Hinkle (2013), but rather than relying on statements of similarities and difference—which may often be dicta—this focuses only on authoritative statements by the Supreme Court (agreeing or disagreeing with a lower court). Frequently the Supreme Court will explicitly note that a particular lower court’s decision is abrogated, even though it did not grant cert in that case. This serves as a counterfactual for how the Supreme Court might have ruled on a different case, had it chosen to review that one instead. This counterfactual parallels a theoretical quantity in the model—the Supreme Court’s treatment of the case it does not review.
4.2 Lower Court Conflict and Supreme Court Review

Conflict between lower courts is one of the best known predictors of *certiorari*; in fact, it is one of the few case attributes explicitly cited as important for *certiorari* in Rule 10 of the Supreme Court’s internal procedures. Scholarship on lower court conflict and its relationship to Supreme Court review is thorough, beginning from studies showing that conflict does indeed predict *certiorari* (see for example Ulmer 1984, Perry 1991, and cites therein). Others identify nuances in this relationship. Estreicher and Sexton (1984) show that legitimate conflict (as opposed to alleged conflict) is very good at predicting *certiorari*; they argue the Supreme Court takes all cases that are “deserving” of review. In a similar vein, Black and Owens (2009) show that “strong” conflict is more likely to lead to review than “weak” conflict. Focusing on when conflict leads to review, Clark and Kastellec (2013) show the Supreme Court’s pattern of reviewing lower court conflict is consistent with an optimal stopping model, in which the Court allows conflict to percolate until it is confident in the correct resolution, then grants cert to resolve the conflict. Like this analysis, their paper claims that conflict is informative. But their model assumes all cases are equally good vehicles for conflict resolution; this analysis explores how and why some cases are more informative than others.

These works have largely focused on conflict as a legalistic or jurisprudential concern. Conflict is assumed to work orthogonally to political motives in predicting review (see e.g. Perry 1991, page 278), arguing justices consider *either* whether they will win on the merits or whether there is conflict in the circuits, but not both). Recently, however, scholars have begun to consider an interactive effect between lower court conflict and ideologically motivated review. As articulated by Perry (1991), “conflict is in the eye of the beholder,” thus, justices can see conflict when it is in their political interest to do so, and be blind
to it when their ideology predisposes them to deny *certiorari*. In other words, conflict may make a case worthy of *cert*, but as Perry’s interviews show, “alone it’s not enough” (Perry 1991, page 246). For example, Black and Owens (2009) argue that policy can amplify or tamper the effects of legalistic concerns, so that when a decision is contrary to a justice’s preferences, conflict will make him much more likely to grant review. Likewise, Epstein, Martin and Segal (2012) show that the relationship between ideology and *cert* votes is only observable when there are some, but not too many, legal factors in favor of granting *cert*.

All these recent works, then, suggest that when the Supreme Court is resolving circuit conflicts, it will not side with the circuit it chooses to review. Even when it is resolving a circuit split, the Supreme Court does not choose which cases to review randomly (Hansford 2011). While no existing work predicts the Court has these incentives, some empirical work suggests the results might hold. In particular, Wasby (2005), Lindquist and Klein (2006), and Summers and Newman (2011) argue the Supreme Court’s reversal rate is lower than it seems. Each study shows that the Supreme Court often mentions other Courts of Appeals decisions in its opinions—decisions it did not choose to review and with which it agrees.

### 4.3 Theory and Hypotheses

The model indicates the Supreme Court will choose to review the cases that will be most informative. Attempting to learn through review affects the Supreme Court’s choice of when to grant *certiorari*, and also affects the Supreme Court’s doctrinal choices. It is not obvious which doctrine the Court will choose to adopt—that which it reviews, or that which it does not review?

Since the Supreme Court is reviewing those cases most likely to be informative, one might imagine the Supreme Court is likely to endorse the doctrine in cases it reviews—since these cases contain more information, one might expect them to be correct. That is, one might
imagine that informative cases are generally supported, if the Supreme Court reviews them simply because it wishes to endorse their content. Or, choosing informative cases may make the Supreme Court more likely to strike down those decisions it reviews, if the Supreme Court is searching for the strongest evidence against its predisposition.

In fact, the model predicts informative cases’ doctrine will not be endorsed, in general. Because the Supreme Court is trying to learn, it should be more likely to strike down the doctrines it chooses to review, and to side with doctrine from the cases it does not review. The model also indicates that this effect is strongest when lower courts are ideologically heterogeneous.

This prediction relies on the Supreme Court’s treatment of both the court it reviews and the court it does not review. Recall the Supreme Court’s choice of doctrine can effectively reverse the decision of a court it does not review. This is because the Supreme Court’s choice of doctrine is binding on all lower courts. If $LC_I$ makes a decision of $A$ and $LC_{II}$ makes a decision of $M$, a ruling of $M$ from the Supreme Court reverses $LC_I$’s decision, regardless of whether the Supreme Court chose to review $LC_I$ or $LC_{II}$’s decision before making her own. Thus, by considering reversals for cases that are not reviewed, the model is able to make predictions about which doctrines are more likely to be struck down—those that are reviewed or those that are not. If the Supreme Court reverses one and only one of the lower courts’ rulings, it must reverse either the case it reviews or the case it does not review. The model predicts the Supreme Court will tend to reverse the case it reviews. This tendency is more pronounced when the lower courts are ideologically heterogeneous than when the lower courts are homogeneous. Figure 4.1 illustrates the parameter space in which the Supreme Court is more likely to reverse the cases it reviews than the cases it does not review. This space is larger when lower courts are ideologically heterogeneous.
Figure 4.1: Parameter space in which taken cases are more likely to be reversed than not-taken cases ($\alpha$ on the horizontal axis and $q$ on the vertical axis). The lighter-shaded area is the space in which this is true only for ideologically heterogeneous lower courts. The darker-shaded area is the space in which this is true for homogeneous and heterogeneous lower courts.

**Doctrinal Hypothesis** The probability of rejecting doctrine in cases that are reviewed is higher than the probability of rejecting doctrine in cases that are not reviewed, especially when lower courts are ideologically heterogeneous.

The extent of ideological heterogeneity in lower courts has profound implications for understanding lower court conflict. When lower courts that are ideologically identical articulate different doctrines, they do so because their information is different. Perry (1991)’s interviews with law clerks about the process of granting *cert* indicate how this might function: one clerk states that, when trying to decide whether or not to review a conflict between circuits, it matters which judges are involved in the conflict: “If it was from a panel that was really undistinguished, there was an attitude that the system would probably cleanse itself.” In other words, if the Supreme Court could be certain that one lower court’s result differed because it has processed information badly (or received incorrect information from lawyers),
then the Court knew the correct outcome and had no need to review to learn. Lower courts
that are ideologically different from one another may also receive different information—but
they also might articulate different doctrines because they process information in a biased
fashion. When Republican appointees in the Fifth and Eleventh circuits differ in their doc-
trines, one expects the conflict to be based on differences in lawyers’ arguments. When
Republican appointees in the Fifth circuit articulate a different doctrine than do Democratic
appointees in the Ninth, one expects the conflict could have arisen for any number of reasons.

4.4 Data

I test this prediction using data on the Supreme Court’s resolution of conflicts in the
Courts of Appeals. I use two previously collected databases of such cases, Lindquist and
Klein (2006) and Summers and Newman (2011). Both databases contain all Supreme Court
cases where the Court states it is resolving a conflict between lower courts, and both identify
all circuit court cases involved in the conflict. Thus, the datasets include Courts of Appeals
cases clustered into conflicts that were resolved by the Supreme Court, so that each conflict
results in one Supreme Court opinion and a number of Courts of Appeals opinions. The
data includes 309 conflicts resolved between 1987 and 1995 (Lindquist and Klein 2006) and
172 conflicts resolved between 2005 and 2010 (Summers and Newman 2011). An average of
4-5 lower courts were involved in each conflict, thus there are 1,345 lower court cases from
the 1987-1995 period and 855 from the 2005-2010 period. The left panel of Figure 4.2 shows
the distribution of the number of lower courts involved in a conflict.

A conflict could, in theory, have many sides, with each circuit articulating a substantially
unique position. In practice, conflicts almost always have two sides, and circuits that par-
ticipate in the conflict espouse one of these two sides. For each Courts of Appeals case, the
datasets record the side of the conflict the lower court chose—side A or side B. These are
Figure 4.2: Histogram of the number of lower courts involved in a conflict resolved by the Supreme Court (left) and distribution of the proportion of circuits rejected (right).

not coded as liberal or conservative sides of the conflict (for example), they simply organize cases into those that supported one side and those that supported the other. Then, for each conflict, the datasets record which of those two sides the Supreme Court supported. When the Supreme Court chose neither of these two sides and announced its own doctrine, this is recorded as well. The Supreme Court’s opinion indicates whether each lower court case’s doctrine was “accepted”—if the Supreme Court agreed with it—or “rejected,” if the Supreme Court did not. If the Supreme Court chose neither of the two sides and wrote its own doctrine, all lower court cases are coded as “rejected.” Notably, this was exceedingly infrequent—in 2202 of 2512 cases, or 88%—the Supreme Court rejects some of the circuits’ doctrine and accepts others’. (In the remaining 12%, the Supreme Court rejected all the doctrines.)

The right panel of Figure 4.2 shows the distribution of the proportion of circuits who had their doctrine rejected in any given conflict. In general, the Supreme Court rejected about
half the circuits and endorsed about half. Finally, each dataset records which of the many lower court cases was reviewed before the Supreme Court made its decision. Thus, an appeals court case is either reviewed or not reviewed and its doctrinal position is either rejected or accepted. Since the Supreme Court’s doctrinal pronouncement is universally binding, it is possible for a lower court case to be not reviewed and rejected.

The model indicates effects should be different when lower courts are homogeneous compared to when they are ideologically heterogeneous. Since 1987, when the data begins, the Supreme Court has been consistently conservative (and has trended more conservative) (Bailey 2007). In the earlier part of this time period—in the data from 1987-1995—the lower courts were also overwhelmingly conservative. Throughout this time period, over half of all panels on the Courts of Appeals had two or three judges appointed by a Republican president. At its peak, in 1993, fully 75% of panels were majority-Republican (Kastellec 2011b). In fact, during much of this time period no circuits were controlled by Democrats. Figure 4.3 shows the number of circuits controlled by Democrats over time. Between 1985 and 1987 almost all circuits were Republican-controlled; between 2005 and 2010 about three circuits—25%—were controlled by Republicans. Therefore, I define the data from 1987-1995 to be an ideologically homogeneous period, and the data from 2005-2010 to be ideologically heterogeneous.

4.5 Analysis

The unit of analysis is a Courts of Appeals case, and the dependent variable is the Supreme Court’s treatment of the doctrine articulated in said case—either it is accepted or it is rejected. The independent variable of interest is whether the case was selected for review, or not. As a referent, Fraction of circuits rejected indicates the fraction of circuits involved whose doctrine was rejected. Consider a conflict with five lower court cases, three of which
Figure 4.3: The number of circuits controlled by Democratic appointees, over time. Between 1987 and 1995 almost no circuits were controlled by Democrats; between 2005 and 2010 more were.

chose doctrine A and two of which chose doctrine B. If the Supreme Court chose doctrine A, Fraction of circuits rejected would be $\frac{2}{5}$; if the Supreme Court chose doctrine B, Fraction of circuits rejected would be $\frac{3}{5}$. If review were random, then in expectation the probability of rejection upon review would be equal to Fraction of cases rejected.

There are a number of relevant control variables that predict whether or not a lower court’s doctrine is adopted. Number of circuits choosing same doctrine counts the number of circuits that agree with the Courts of Appeals case in question, and thus measures the popularity of a given doctrine. Consider again a conflict with five lower court cases, three of which chose doctrine A. For each of those three cases, Number of circuits choosing same doctrine is 3. For the two circuits choosing doctrine B, Number of circuits choosing same doctrine is 2. This measure of popularity has been shown to be a good predictor of Supreme Court treatment (Lindquist, Haire and Songer 2007).

An ideological model might predict that the probability of doctrinal rejection rises overall
when lower courts become more ideologically heterogeneous—there are simply more decisions with which the Supreme Court disagrees, and therefore there is just more to strike down (see e.g. Lindquist, Haire and Songer 2007). Indeed, during the ideologically heterogeneous period of the data, the average proportion of reversals was .69; during the homogeneous period, the average proportion of reversals was .48. Therefore, it is important to control for the ideology of the court espousing the given doctrine, to ensure the Supreme Court is not simply reviewing and reversing based on ideological factors. To control for this, I account for the proportion of Democratic appointees on the circuit from which the decision arose. The circuit is a better measure than the identity of judges on the panel when considering doctrinal articulation: because of horizontal *stare decisis* within circuits (Kornhauser 1995), even judges who are unrepresentative of circuit ideology might reflect the ideological preferences of the circuit on the whole.

### 4.6 Results

To explore the relationship between review and reversal, I begin with a simple look at the data: comparing the chances a doctrine is rejected upon review to the overall chance it is rejected. Table 4.1 presents the results from logistic regressions, where the dependent variable is whether a circuit’s doctrine was rejected or not. As predicted, the coefficient on *Review* is positive and statistically significant during the heterogeneous period, but indistinguishable from 0 during the period with homogeneous lower courts. In 2005-2010, the probability of rejecting doctrine is much higher for reviewed than unreviewed cases. The third column of Table 4.1 presents a regression run on both periods. The coefficient on *Review* is indistinguishable from zero—during the homogeneous period, the Supreme Court

---

3 I alternatively included an indicator for the circuit being majority-Democrat; results are identical.

4 There is an ecological inference problem that arises from such a choice, but for this to confound the results it would have to be the case that Republican judges on Democratic circuits are more conservative than their counterparts in Republican circuits (and vice versa).
was as likely to strike down doctrines it reviewed as doctrines it did not review. Likewise, the coefficient on lower court heterogeneity is indistinguishable from zero—among cases that were not reviewed, probability of rejection is equally high during both time periods. The coefficient on the interaction term, however, is large, positive, and statistically significant—during the heterogeneous period, the Supreme Court was noticeably more likely to strike down those doctrines it reviewed.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Case is Reviewed</td>
<td>0.12</td>
<td>0.89*</td>
<td>0.13</td>
<td>0.14</td>
<td>0.20</td>
<td>0.14</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.20)</td>
<td>(0.14)</td>
<td>(0.11)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>-0.13</td>
<td></td>
<td></td>
<td>0.74*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td></td>
<td></td>
<td>(0.24)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneous × Reviewed</td>
<td>4.63*</td>
<td>5.34*</td>
<td>4.95*</td>
<td>4.63</td>
<td>5.34</td>
<td>4.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.40)</td>
<td>(0.26)</td>
<td>(0.35)</td>
<td>(0.40)</td>
<td>(0.26)</td>
<td></td>
</tr>
<tr>
<td>Fraction of Circuits Rejected</td>
<td>-2.35*</td>
<td>-2.83*</td>
<td>-2.51*</td>
<td>-2.35</td>
<td>-2.83</td>
<td>-2.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.22)</td>
<td>(0.15)</td>
<td>(0.18)</td>
<td>(0.22)</td>
<td>(0.15)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1345</td>
<td>855</td>
<td>2200</td>
<td>1345</td>
<td>855</td>
<td>2200</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: The probability of rejection, predicted by fraction of circuits rejected. Standard errors in parentheses. * indicates $p < .05$.

The top row of Figure 4.4 illustrates this graphically. The figure plots the predicted probability of rejection for any given case, against the proportion of doctrines rejected in that conflict. Consider a conflict with five circuits involved. If three of these circuits had their doctrine rejected, the probability of rejection upon review would be 3/5 if review were random. If that were the case, the predicted probability of rejection would equal the proportion of doctrines rejected for each observation. Thus, if there were no relationship between probability of doctrinal rejection and *certiorari*, the data would fall precisely on the diagonal. If review and rejection were not independent, the prediction for *reviewed* cases
and unreviewed cases would diverge from the diagonal.

As shown in the top left panel, in the period of lower court homogeneity (1987-1995) the Supreme Court was no more likely to reject the doctrines it reviewed than those it did not review: the proportion of cases rejected falls cleanly on the 45-degree line. In other words, the Supreme Court’s review decisions and the Supreme Court’s doctrinal choices are independent. In contrast, in the heterogeneous period (2005-2010), the proportion rejected underpredicts the probability of rejection for reviewed cases and overpredicts rejection for cases not reviewed. The top right panel shows that the probability for reviewed cases lies above the 45-degree line: they are rejected more than they would be if review were random.

To further illustrate the magnitude of this phenomenon, it is useful to change perspective and ask which case is most likely to have its doctrine rejected. As previously discussed, circuit conflicts are often protracted so that many circuits weigh in on the issue; a conflict typically lasts about five cases. In the period of lower court heterogeneity, when the Supreme Court rejected the doctrine of only one of the many cases below, there was a 68% chance it was the case it took. Were the Supreme Court reviewing at random, there would be a 20% chance of this occurring.

It is clear, then, that when lower courts are ideologically heterogeneous the Supreme Court becomes more likely to reject the doctrines it chooses to review. Table 4.2 shows this continues to hold, even controlling for how many circuits chose the same doctrine, and the ideology of the court in question. During periods of ideological heterogeneity, a purely ideological theory would predict the Supreme Court would target biased, ideologically distant courts for review and reversal. But the learning model predicts that when some lower courts are biased, the Supreme Court will occasionally prefer to review and reverse the decisions of ideologically proximate lower courts. The results support this hypothesis: even controlling for ideological composition of the lower courts, the Supreme Court’s doctrinal choices are
Figure 4.4: Probability of rejecting doctrine in cases granted review and cases not granted review, by proportion of cases whose doctrine is rejected (top row). The left plot shows data from when the Courts of Appeals were largely homogeneous; the right plot shows data from when they were more ideologically heterogeneous.
<table>
<thead>
<tr>
<th></th>
<th>1987-1995 Homogeneous</th>
<th>2005-2010 Heterogeneous</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case is reviewed</td>
<td>0.03</td>
<td><strong>0.84</strong></td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.16)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Number of circuits choosing same doctrine</td>
<td><strong>-0.15</strong></td>
<td><strong>-0.14</strong>*</td>
<td><strong>-0.14</strong>*</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Democratic circuit</td>
<td>0.34</td>
<td>0.16</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td>(0.50)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>Heterogeneous × reviewed</td>
<td></td>
<td><strong>0.80</strong>*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.19)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.22</td>
<td>0.35</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.25)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>N</td>
<td>1345</td>
<td>855</td>
<td>2200</td>
</tr>
</tbody>
</table>

Table 4.2: The probability of rejection, controlling for number of circuits choosing same doctrine and ideology of lower court. Coefficients from logistic regressions. Standard errors in parentheses. * indicates $p < .05$.

still consistent with the learning model. The coefficient on the interaction term is still positive and statistically significant, indicating that when lower courts are heterogeneous the Supreme Court is more likely to reject the doctrines it chooses to review. In the ideologically heterogeneous period, 2005-2010, the Supreme Court was much more likely to reject doctrine in cases it reviewed; in the homogeneous period, 1987-1995, the Supreme Court was about as likely to reject doctrine in cases it reviewed as cases it did not review. Additionally, it is interesting to note that the coefficient on Number of circuits choosing same doctrine is negative and statistically significant in all three models. This replicates the result shown in Lindquist and Klein (2006), namely, that the Supreme Court tends to agree with the majority of circuits. In other words, the more circuits that agree A is correct, the more likely is the Supreme Court to agree that A is correct.

Figure 4.5 presents a graphical representation of the goodness-of-fit of the logistic model.

72
Figure 4.5: Graphical depiction of goodness-of-fit of Combined logistic regression. See text for details.

The horizontal axis shows the predicted probability of doctrinal rejection, as fitted by the model; the vertical axis indicates whether or not the doctrine in question was in fact rejected. Points that fall to the top left and bottom right of the plot are correctly predicted; those on the top left and bottom right are incorrectly predicted. The lowess line (span=.6) shows how well the predicted probability correlates with observed reversal. The model performs well, but not strikingly so. Of those cases with a very low probability of rejection (less than .25), none were rejected; of those with a relatively high predicted probability of rejection (greater than .71), 68% were rejected. Thus, review is important for understanding the Supreme Court’s choice of doctrine. The Supreme Court is much more likely to grant certiorari in those cases it ultimately reverses.

4.6.1 Strategic litigants

Not all cases decided in the Courts of Appeals petition for certiorari, however. Only 20% of cases terminated on the merits do [Administrative Office of the United States Courts...](#)
There are two ways by which this could influence the results shown above. First, if cases rose to the Supreme Court infrequently, the Court’s choices would be constrained by available cases. If the Supreme Court wished to resolve a conflict and had only one cert petition available for resolution, its review decision would not be strategic.

This first concern is not a binding constraint on the Supreme Court, empirically. The Supreme Court often chooses to hold petitions for cert over from one term to the next, which assuages some of the constraint. Also, often both sides of the conflict choose to petition for review in the same year. Of the 1,081 “conflict-years” in the Lindquist and Klein data, 174, or 16%, allow the Supreme Court to hear either side of the conflict. Interestingly, and consistent with the predictions of 2.2.3, the Supreme Court typically does not choose to hear both sides of the conflict. Of the 174 opportunities to consolidate, in 134 chose not to hear all the cases that had petitioned.

Second, the Supreme Court’s review could be constrained by the sample of cases in which cert was requested. Petitioning for certiorari is an expensive and demanding process. Therefore, litigants are likely to be selective in their petitions for review—specifically, winning litigants may never petition for cert and litigants who lose in lower courts may be more likely to petition for cert only in those cases where they believe they are likely to win at the Supreme Court (Cameron and Kornhauser 2005, McGuire et al. 2009). If litigants are more likely to petition for certiorari on cases that are more likely to be reversed, it could be litigant behavior that explains the results seen above, rather than strategic behavior by the justices.

To see if this constrained the Court, I randomly selected a subset of cases from each of the time periods, and coded whether or not litigants had petitioned for certiorari in each of the lower court cases. The results show that petitions are no more likely in cases that end

---

5I coded this by looking the cases up in Lexis-Nexis and Westlaw, and checking the subsequent history.
### Table 4.3: Whether litigants requested *certiorari*, and whether the Supreme Court ultimately agreed or disagreed with the doctrine articulated by the lower court.

<table>
<thead>
<tr>
<th></th>
<th>Heterogeneous</th>
<th>Homogeneous</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reversed</td>
<td>Affirmed</td>
</tr>
<tr>
<td>Not Requested</td>
<td>81</td>
<td>61</td>
</tr>
<tr>
<td>Requested</td>
<td>59</td>
<td>56</td>
</tr>
</tbody>
</table>

The probability of a *cert* request is slightly higher among cases that were ultimately affirmed, suggesting the Supreme Court’s likelihood to reverse cases actually reviewed is not driving by litigant selection.

Table 4.3 shows the distribution of *certiorari* petition by ultimate disposition, in the subset of cases studied. The probability of petitioning for *certiorari* among cases that were ultimately reversed is 42% and 38%; the probability of petitioning for *cert* among cases that were ultimately affirmed is 48% and 46%. Thus, the probability of petitioning for *certiorari* is, if anything, slightly higher among cases that are actually affirmed. This suggests it is the Supreme Court’s choices that drive the result, and not litigant strategy. It also suggests there is room for growth in understanding litigant strategy, which is discussed below.

### 4.7 Discussion and Conclusion

Not all cases are equally good vehicles for resolving conflicts. The best vehicles—those cases that are most informative about what doctrine to choose—often articulate doctrine that is ultimately rejected, especially when the lower courts are heterogeneous. By analyzing the Supreme Court’s resolution of lower-court conflicts, and studying the relationship between the cases it chooses to review and the doctrines it chooses to support, this chapter illustrates one effect of the Supreme Court’s attempts to learn from the Courts of Appeals. In particular, the fact that review is not random reinforces the conclusion that the Supreme Court’s primary goal is informing its doctrinal decisions, and reveals that aggregation of lower court decisions and direct history respectively.
has implications beyond the review stage.

It has long been acknowledged that uncertainty plays a large role in explaining Supreme Court decision making, but for the most part this uncertainty has been construed as pertaining to case facts. But there is a coincident source of uncertainty for the court, with more far-reaching implications: the uncertainty of how to craft doctrine for resolving large swaths of disputes, rather than individual cases. While the Supreme Court may also be unsure about the facts of a particular case, here the Court’s focus is taken to be a broader question. To answer this broader question, the Supreme Court learns from a group of lower courts’ decisions.

Learning from a group of lower courts means the Supreme Court’s decisions are necessarily contextual. The decision of whether to review any given case is premised on outcomes of other cases; upon review, doctrinal decisions are influenced by many lower courts, not just the one whose decision is being reviewed. Based on this theory, the model identifies which cases should be reviewed, and which doctrines should be chosen. By relying on the resolution of conflicts between the lower courts, and the Supreme Court’s own statement of which circuits’ positions are being supported and which not, the results here show how the Supreme Court’s doctrinal choices coincide with its review decisions.

The results also reinforce the conclusion that jurisprudential and political concerns are complementary. Often, lower court conflict and lower court non-compliance are seen as substitutable reasons for granting cert. The former is construed as a concern for uniformity in the law and legalistic motives; it is exemplary of the jurisprudential mode identified by Perry (1991). The latter is seen as arising from political concerns, and as exemplifying the outcome mode of review. This analysis illustrates how these forces can work together. Even when the Supreme Court states it is reviewing a case in order to resolve a conflict between lower courts, it is still more likely to do so by reviewing those cases whose doctrine it rejects.
There are a number of particularly interesting possibilities for future work. First, one might consider lower court conflicts absent Supreme Court intervention. Some conflicts are never reviewed; even among those that are, percolation and evolution predates and postdates Supreme Court intervention. It would be useful to trace lower court conflicts without relying on the Supreme Court’s distillation of the conflicts. Second, it would be useful to further understand litigant behavior, and particularly to focus on conflicts and litigant strategy as dynamic quantities. Clark and Kastellec (2013) show that the Supreme Court is more likely to review “late” conflicts—issues that have been percolating and have been addressed by many courts—than “early” conflicts that have not percolated sufficiently. One might consider whether litigants, cognizant of the Supreme Court’s preferences about review, are more likely to petition at some times than others.
Chapter 5

Reversal

I turn now to the most straightforward quantity that results from Supreme Court decision-making—whether the lower court’s decision is affirmed or reversed. When the Supreme Court is deciding on the constitutionality of a major law, as in NFIB v. Sebelius, or when the Supreme Court’s disposition has a direct impact on election outcomes, as in Bush v. Gore, this disposition is an object of great substantive interest. It is also the object that judicial politics scholars study most. For example, the dispositional vote is used in the standard measures of Supreme Court ideology (Martin and Quinn 2002) and is the dependent variable in a number of seminal works in judicial politics (George and Epstein 1992, Stimson, Mackuen and Erikson 1995, Segal and Spaeth 2002, Richards and Kritzer 2002).

Many theories have evolved to explain how the Supreme Court will resolve cases (Cameron, Segal and Songer 2000, Beim, Hirsch and Kastelles 2012) from a hierarchical perspective. Few, however, have empirically evaluated these hierarchical theories based on the Supreme Court’s decisions. Most theories of the hierarchy have analyzed decisions of which cases to review. (A notable exception is Carrubba and Clark 2012, discussed in more detail below.) This chapter analyzes the Supreme Court’s dispositional votes—which decisions it chooses
to reverse, and which to affirm—in light of the theory of hierarchical learning.

### 5.1 Understanding Reversals

A large body of literature on the judicial hierarchy claims the Supreme Court’s review decisions are governed by a desire to overturn decisions with which they disagree. In the move to understand which cases the Supreme Court will reverse, many have focused on predicting dispositions based on lower court judges’ ideology and decisions. These theories have been tested amply, and the theory is successful at explaining much of the Supreme Court’s *certiorari* decisions. For example, it is well-known that conservative courts are more likely to review liberal decisions than conservative decisions (Cameron, Segal and Songer 2000; Beim, Hirsch and Kastellec 2012). Caldeira, Wright and Zorn (1999) show *certiorari* votes during the 1982 term were “strategic” in the sense that justices looked forward to votes on the merits, and Black and Owens (2012a) show the Rehnquist court was much more likely to review liberal lower courts’ liberal decisions than other decisions and show that this is particularly true for conservative justices. These results have begun to receive criticism (see e.g. Klein and Hume 2003; Carrubba and Clark 2012), but on the whole, the Supreme Court is known to target ideologically distant courts for review and is thought to target ideologically suspicious decisions. Intuitively, many argue that they do so because they wish to reverse these decisions.

It is natural, then, to ask whether in fact the Supreme Court *does* reverse these decisions. Surprisingly, this question has received very little attention. The few studies that have investigated the rate at which the Supreme Court reverses lower courts looks only at aggregate-level statistics, considering which *circuits* are most likely to be reversed, rather than which *panels* are most likely to be reversed (Scott 2006; Posner 2000). These papers find that the Supreme Court is more likely to review circuits distant from it—for example, as
is well-known in scholarly and popular work, the Rehnquist court targeted the Ninth Circuit for review and reversal \cite{Wasby2005}. But there is an ecological inference problem arising from this method: assuming that Republican judges in liberal circuits will be reversed at the same rate as their liberal colleagues. This chapter builds primarily on the results of two prior studies of hierarchically-motivated reversals that focus on the lower court judges who made the decision. \cite{Hall2009} considers whether liberal lower courts are more likely to be reversed, taking an experimental perspective that relies on the random assignment of judges to Courts of Appeals panels. \cite{ClarkCarrubba2012} show that the probability of reversal is highest for those lower courts who are distant, but not too distant, from the Supreme Court. Both of these consider only the ideology of the lower court, not the decision the lower court makes; I consider both factors.

Even from this literature, some evidence is inconsistent with existing theories. For example, because the Supreme Court is assumed to review cases in order to reverse them; affirmances may occur, occasionally, but only they are mistakes. Affirming the decision of a lower court happens only as an accident of incomplete information. In the model presented here, however, an affirmation is doctrinally useful. This alternative intuition might better explain the Supreme Court’s opinions affirming decisions of lower courts. Since affirmances are mistakes in disciplinary models, so they can be assumed to yield short, curt opinions without much argumentation, but here they are equally effective vehicles for communicating doctrine\footnote{Unlike in Supreme Court review, when Courts of Appeals review district courts’ decisions law making is largely a secondary concern. There, publication is optional, so a panel may abstain from writing a precedential opinion if it so chooses. \cite{MerrittBrudney2001} show that reversals are much more likely to be published. Indeed, this is largely institutionally-driven: many circuits suggest reversals should be published. \cite{EpsteinLandesPosner2013} also consider the doctrinal import of opinions affirming the court below, noting they are often easier to craft: “effort aversion should predispose judges to affirm. It is easier to write an opinion affirming than one reversing, because when affirming one can rely heavily on the lower court’s decision whereas reversal requires refuting the grounds of that decision.” These patterns corroborate the claim that auditors’ affirmances are not as elaborate as their reversals.}
The theories imply those circuits most likely to be reviewed are also those most likely to be reversed. But, as Hofer (2010) shows, that is not the case. Figure 5.1 shows the relationship between review and reversal, by circuit. For each circuit-year between 1999 and 2008, the horizontal axis shows the number of cases the Supreme Court reviewed from that circuit, divided by the number of cases that circuit disposed of on the merits in that year. The vertical axis shows the number of cases the Supreme Court reversed from that circuit in that year, divided by the total number of cases it heard from that circuit in that year. If the Supreme Court were reviewing for the purpose of reversal, one would expect the proportion of cases reversed to rise along with the proportion of cases reviewed. However, this is not the case: the relationship is fairly flat.

Figure 5.1: Statistics from Hofer (2010) on the proportion of cases of reviewed (horizontal axis) and proportion of cases reversed or vacated (vertical axis), and a lowess line. There is no noticeable positive association between the two quantities.
5.2 Theory and Hypotheses

From the model presented in Chapter 2, there are two primary predictions pertaining to the Supreme Court’s dispositional choices. The first arises because in the learning model, moderate decisions are reviewed for informational purposes. Purely liberal and conservative decisions, in contrast, are reviewed only for purposes of increasing doctrinal uniformity. Therefore, whenever the Supreme Court reviews a decision of A or B in the learning model, it should reverse the decision.

Extremity Hypothesis: Upon review, liberal and conservative decisions should be reversed more often than mixed decisions.

The second prediction arises in the version of the model with ideological heterogeneity. Recall that in the model, unbiased lower courts make all decisions—A, B, and M, while in equilibrium biased lower courts only make decisions of A and M. When the unbiased lower court makes a decision of B and the biased lower court makes a decision of A, the Supreme Court may review the unbiased lower court’s decision of B. In fact this is the only circumstance under which the unbiased lower court’s decision of B will be reviewed, and these decisions will always be reversed. When the lower courts make decisions (M, A), the Supreme Court will choose to review LC’s decision of M. This is likewise the only circumstance under which the Supreme Court would review the unbiased lower court’s decision of M. These M decisions will sometimes be reversed, and sometimes be affirmed. Thus, the Supreme Court is more likely to reverse unbiased lower courts’ moderate decisions than their decisions of B.

A parallel result arises for the biased lower court. Biased lower courts’ decisions of A will sometimes be reviewed when the lower courts make decisions (B, A). This is the only instance in which the biased lower courts’ decision of A will be reviewed, and these decisions
will always be reversed. Biased lower courts’ decisions of $M$ may be reviewed regardless of which decision the unbiased lower court makes (whether the courts make decisions $(A,M)$; $(B,M)$; or $(M,M)$). These decisions of $M$ will sometimes be reversed, and sometimes be affirmed. As a result, biased lower courts’ decisions of $A$ will *always* be reversed, while biased lower courts’ decisions of $M$ will sometimes be reversed and sometimes be affirmed.

Suppose the Supreme Court is conservative, so that conservative lower courts are “unbiased” and liberal lower courts are “biased.” It is then natural to think of decisions of $B$ decisions as conservative and decisions of $A$ as liberal. Thus, a biased lower court—a liberal lower court—will make liberal and moderate decisions only. When that liberal lower court makes a liberal decision, it will be more likely to be reversed than when the liberal lower court makes a moderate decision. In summary, the prediction is as follows:

**Opposite Decision Hypothesis:** Conservative lower courts’ liberal decisions will be more likely to be reversed than conservative courts’ moderate decisions. Liberal lower courts’ liberal decisions will be more likely to be reversed than liberal lower courts’ moderate decisions.

### 5.3 Data

To test these hypotheses, I turn to the decisions made by the Burger Court between 1970 and 1986. During this time the Court decided about 1,174 cases that arose from the Courts of Appeals, for an average of 73 cases a year. (The Court also decided cases that came from state courts, and those of original jurisdiction that had not previously been heard by any other court, but these are omitted from the analysis. Consolidated cases are also excluded.) These cases are the product of the data presented in Chapter 3; these are the cases the Burger Court chose to review. Now, the dependent variable is the Supreme Court’s disposition in the case. I consider only cases that were decided on the merits with a full written opinion of
the Court—memoranda are excluded. A case is Affirmed if and only if the Supreme Court affirms the decision in whole. This parallels the model’s prediction of whether the Supreme Court’s chosen doctrine agrees or disagrees with the doctrine articulated by the lower court it is reviewing.

<table>
<thead>
<tr>
<th>Decision Description</th>
<th>Liberal</th>
<th>Conservative</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 + remanded</td>
<td>10</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Affirmed</td>
<td>167</td>
<td>129</td>
<td>75</td>
</tr>
<tr>
<td>Affirmed and reversed (or vacated) in part</td>
<td>4</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Certification to a lower court</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Petition denied or appeal dismissed</td>
<td>8</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Reversed</td>
<td>162</td>
<td>91</td>
<td>66</td>
</tr>
<tr>
<td>Reversed and remanded</td>
<td>139</td>
<td>88</td>
<td>72</td>
</tr>
<tr>
<td>Stay, petition, or motion granted</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vacated</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Vacated and remanded</td>
<td>52</td>
<td>36</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 5.1: Distribution of lower court decision (Liberal, Conservative, or Mixed) and Supreme Court disposition.

Decisions that are affirmed in part and reversed (or vacated) in part are coded as not affirmed. This happens in 1.3% of cases overall, and 4.2% of cases that received mixed outcomes in the lower court. Although the hypothesis is framed as which cases will be reversed, in the empirical specification I use Affirmed rather than Reversed. I do this for clarity in discussion of the results, since when the Supreme Court chooses not to affirm a decision, it may either reverse the decision, remand it, vacate it, or otherwise Not Affirm it. As a matter of jurisprudence, these differences are meaningful, but from the perspective of the model they are not; therefore, I use only Affirmed and Not Affirmed. Table 5.1 shows the distribution of Supreme Court disposition by lower court outcome.

To test the dispositional hypotheses described above, I include measures of Supreme Court ideology, lower court ideology, and the direction of the lower court’s decision as in-
dependent variables. To measure the direction of the lower court’s decision, I use the same measure described in Chapter 3. I measure the direction of the lower court’s decision by classifying parties into liberal and conservative, and noting which prevailed. In general, “underdogs,” defendants in criminal trials, complainants bringing suits against their employers, and similarly positioned individuals are considered “liberal.” When they win, the outcome is considered a liberal one. When conservative parties win on all counts, the outcome is considered conservative. When each party wins on some counts, the outcome is coded as Mixed. See Chapter 3 for more information.

To measure the ideology of the Supreme Court and the Courts of Appeals, I use both the same approach as in Chapter 3 and an additional measurement strategy. The Burger Court was largely conservative Blasi (1983) and oversaw a largely conservative group of Courts of Appeals (see Chapter 4). As a result, it is possible to use a straightforward measure of judicial ideology: as in Chapter 3 I consider the Supreme Court to be conservative and measure lower courts’ ideology based on the party of the appointing president. Again, panels composed of three Democratic nominees are denoted DDD and are considered “biased,” since they are ideologically distant from the Supreme Court; panels of three Republican nominees are denoted RRR and are considered “unbiased.”

I also present results using Judicial Common Space scores Epstein et al. (2007). The scores allow the Supreme Court’s ideology to change over time, and also capture heterogeneity between judges of the same party. For Courts of Appeals judges, the scores are based on nominations and appointments, but are more refined than the party of the appointing president. A Courts of Appeals judge’s score is equal to the NOMINATE score Poole and Rosenthal (1997) of the president who appointed him, or the NOMINATE score of the judge’s home-state senator if the senator is of the same party as the president. This is because of the practice of senatorial courtesy, under which the president consults with senators be-
fore making nominations in their state. For the Supreme Court justices, the scores are a
monotonic transformation of Martin and Quinn scores [Martin and Quinn 2002], which are
dynamic judicial ideal points derived from the justices’ dispositional votes. Figure 5.2 shows
the median Judicial Common Space score by panel type, illustrating the differences between
these two measurement strategies.

Figure 5.2: Median Judicial Common Space Score by panel composition.

5.4 Analysis

Overall, about 29% of cases in the data are affirmed. As Figure 5.3 shows, this has been
fairly constant over time in the modern era of the Supreme Court—notably, this includes
times when the lower courts are more and less aligned with the Supreme Court. The rate
of affirmance, however, has not changed noticeably and has remained somewhat high, which
again suggests the Supreme Court’s goal is not reversing decisions with which it disagrees.

Table 5.2 shows the rates of affirmance by panel type and decision direction. There are
no striking differences between the cells. The purely conservative RRR panels are not most
likely to be affirmed. Lower courts’ liberal decisions seem as likely to be affirmed as their conservative decisions—and perhaps slightly more likely to be affirmed. Similarly, the more extreme Conservative and Liberal decisions are *not* more likely to be reversed than Mixed decisions.

<table>
<thead>
<tr>
<th></th>
<th>Conservative</th>
<th>Mixed</th>
<th>Liberal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDD</td>
<td>22.35</td>
<td>22.64</td>
<td>26.32</td>
<td>23.93</td>
</tr>
<tr>
<td>DDR</td>
<td>27.17</td>
<td>25.62</td>
<td>28.67</td>
<td>27.38</td>
</tr>
<tr>
<td>RRD</td>
<td>32.27</td>
<td>28.45</td>
<td>34.06</td>
<td>32.04</td>
</tr>
<tr>
<td>RRR</td>
<td>26.15</td>
<td>36.59</td>
<td>31.71</td>
<td>29.53</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>28.00</strong></td>
<td><strong>27.49</strong></td>
<td><strong>30.51</strong></td>
<td><strong>28.76</strong></td>
</tr>
</tbody>
</table>

Table 5.2: Probability of affirmance by panel type and decision direction.

These results are replicated in a logistic model, Model (1), the results of which are presented in the leftmost column of Table 5.3. The model includes decision direction and
lower court composition as predictors of affirmance by the Supreme Court. The results mirror the findings shown above: there are no significant differences across panel type or decision direction. The Supreme Court’s reversal behavior does not seem to be predicted by the ideology of lower court judges or the decisions they make. Holding panel composition fixed at its mode, DDR, the predicted probability of a liberal decision being affirmed is .29, the predicted probability of a conservative decision being affirmed is .36, and the predicted probability of a mixed decision being affirmed is .27. Thus, there is no support for the Extremity hypothesis, which predicts moderate decisions should be more likely to be affirmed than either conservative or liberal decisions. Under this specification there is also no support for the Opposite Decision hypothesis.

The insignificance of lower court ideology is also noteworthy. The Supreme Court is no more likely to reverse some panels than others. This finding stands in contrast to studies that find the Supreme Court is more likely to reverse the decisions of ideologically distant circuits, such as Posner (2000) and Scott (2006). More nuanced measures of lower court ideology, and accounting for the direction of the lower court’s decision, cause that finding to disappear.

Furthermore, the results shed light on the findings of Hall (2009). He shows that lower courts composed of three Democratic appointees are far more likely to be reviewed and reversed, as compared to not reviewed or affirmed. Model (1) suggests that finding is a result of the Supreme Court being more likely to review ideologically distant panels, not to reverse them.

To more fully understand the role of ideology, I restrict the sample to cases where the Burger court was most conservative. Although the Burger court was considered quite conservative overall, it was particularly so in the area of criminal law. Therefore, Model (2) con-

\[2\] Predicted probabilities calculated via simulation in Zelig (Carnes 2007).
siders only criminal cases, which should increase accuracy in ideological estimation. These results are presented in the righthand column of table 5.3.

<table>
<thead>
<tr>
<th></th>
<th>(1) All Cases</th>
<th>(2) Criminal Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (Liberal, DDD)</td>
<td>-1.12*</td>
<td>-2.48*</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.74)</td>
</tr>
<tr>
<td>Conservative</td>
<td>-0.16</td>
<td>2.89*</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td>(0.98)</td>
</tr>
<tr>
<td>Mixed</td>
<td>-0.12</td>
<td>2.48*</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
<td>(1.10)</td>
</tr>
<tr>
<td>DDR</td>
<td>0.25</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.81)</td>
</tr>
<tr>
<td>RRD</td>
<td>0.47</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.86)</td>
</tr>
<tr>
<td>RRR</td>
<td>0.40</td>
<td>1.57</td>
</tr>
<tr>
<td></td>
<td>(0.37)</td>
<td>(0.94)</td>
</tr>
<tr>
<td>Conservative × DDR</td>
<td>0.41</td>
<td>-0.68</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(1.10)</td>
</tr>
<tr>
<td>Mixed × DDR</td>
<td>-0.03</td>
<td>-3.22*</td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
<td>(1.56)</td>
</tr>
<tr>
<td>Conservative × RRD</td>
<td>0.50</td>
<td>-1.16</td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
<td>(1.15)</td>
</tr>
<tr>
<td>Mixed × RRD</td>
<td>-0.08</td>
<td>-2.14</td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
<td>(1.35)</td>
</tr>
<tr>
<td>Conservative × RRR</td>
<td>0.27</td>
<td>-2.43↑</td>
</tr>
<tr>
<td></td>
<td>(0.57)</td>
<td>(1.24)</td>
</tr>
<tr>
<td>Mixed × RRR</td>
<td>0.20</td>
<td>-1.57</td>
</tr>
<tr>
<td></td>
<td>(0.64)</td>
<td>(1.60)</td>
</tr>
<tr>
<td>N</td>
<td>1174</td>
<td>261</td>
</tr>
</tbody>
</table>

Table 5.3: Probability of affirmance, Burger years. Standard errors in parentheses. * indicates $p < .05$; † indicates $p = .051$.

Considering the coefficients themselves, the Supreme Court’s conservative position is now evident. The coefficients on both Conservative and Mixed decisions are large, positive, and statistically significant, indicating that liberal decisions are much less likely to be affirmed.
than conservative decisions or mixed decisions. (However, note the counter-intuitive negative coefficient on Republican panels’ conservative decisions, showing these cases are less likely to be affirmed.)

While this is unsurprising in an ideological framework, it is important to note that existing formal models do not predict such a result. Instead, they predict that the Supreme Court will only review liberal decisions, and will be more likely to reverse them the more distant the lower court that made them. To further explore the relationship between lower court ideology and reversal, Table 5.4 presents the results of logistic regressions using Judicial Common Space scores instead of presidential appointment [Epstein et al. (2007)]. Distance measures the absolute distance between the median justice on the Supreme Court and the median judge on the Courts of Appeals panel. Model (3) parallels the results shown above—again, the Supreme Court is no more likely to affirm liberal, conservative, or mixed decisions. Holding distance constant, there is no support for either the opposite decision hypothesis or the extremity hypotheses.

Model (4) considers only criminal cases. The results show no support for either hypothesis, but again, considering only criminal cases does reveal clear ideological patterns. In particular, conservative decisions are much more likely to be affirmed, particularly when they are made by distant lower courts. Even under this specification, holding distance constant there is no increase in probability of affirming mixed as opposed to liberal decisions. Thus there is again no support for either the extremity or the opposite decision hypotheses.

Ideology and decision direction may interact in a more nuanced fashion, however. Spatial models of decision-making would suggest that a distantly conservative lower court who makes a conservative decision, and a distantly liberal lower court who makes a conservative decision, should be seen differently. Model (5) accounts for this. When the Supreme Court is more conservative than the lower court, Conservative Lower Court takes on a value of 0. When
the Supreme Court is more liberal than the lower court, Conservative Lower Court takes on a value of 1.

<table>
<thead>
<tr>
<th></th>
<th>All Cases</th>
<th>Criminal Cases</th>
<th>All Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (Liberal)</td>
<td>-0.71*</td>
<td>-1.40*</td>
<td>-0.39</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.29)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.39</td>
<td>-0.97</td>
<td>-1.27</td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td>(0.90)</td>
<td>(0.72)</td>
</tr>
<tr>
<td>Conservative</td>
<td>0.14</td>
<td>1.15*</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.41)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>Mixed</td>
<td>-0.09</td>
<td>0.45</td>
<td>-0.36</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.58)</td>
<td>(0.44)</td>
</tr>
<tr>
<td>Conservative × Distance</td>
<td>0.12</td>
<td>2.61*</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(0.58)</td>
<td>(1.29)</td>
<td>(1.11)</td>
</tr>
<tr>
<td>Mixed × Distance</td>
<td>0.11</td>
<td>2.39</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>(0.69)</td>
<td>(1.85)</td>
<td>(1.28)</td>
</tr>
<tr>
<td>Conservative Lower Court</td>
<td>-0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservative × Conservative LC</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed × Conservative LC</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservative LC × Distance</td>
<td>0.64</td>
<td></td>
<td>(1.60)</td>
</tr>
<tr>
<td>Conservative × Conservative Lower Court × Distance</td>
<td>0.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed × Conservative</td>
<td>2.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.99)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N 822 187 822

Table 5.4: Probability of affirmance, Burger years, using median Judicial Common Space score for lower courts and Supreme Court median Judicial Common Space score. * indicates p < .05. Standard errors in parentheses.

This model tests the formal ideological prediction yielded by [Cameron, Segal and Songer (2000)]—namely, that when the lower court is more liberal than the Supreme Court, and
when the lower court makes a liberal decision, the more liberal the lower court is, the less likely the Supreme Court will be to affirm the decision. The coefficient on Distance evaluates this hypothesis; it is significant at \( p = .076 \). Thus, there is slight support for the disciplinary theory, but not conclusive support. The result calls out for concrete theoretical explanations as to why this might occur. The model continues not to reveal support for either the opposite decision or the extremity hypotheses, reinforcing the inconclusive nature of the study of dispositions.

### 5.5 Discussion and Conclusion

The Supreme Court’s review decisions have been shown to be consistent with both ideologically-motivated theories of discipline (see for example [Cameron, Segal and Songer](2000) and [Clark (2009)]), and with the theory presented in Chapter 2 that claims the Supreme Court learns from lower courts. Few studies have investigated whether the Supreme Court’s decisions on the merits are consistent with hierarchical theories. This chapter considers two hypotheses generated by a model of learning—the opposite decision hypothesis and the extremity hypothesis. Neither of these receive dispositive support.

It is noteworthy, also, that the results also fail to reveal any support for the predictions of purely ideological theories. The model presented in [Cameron, Segal and Songer (2000)] predicts that among lower courts’ liberal decisions, those made by more liberal courts will be more likely to be reversed than those made by conservative courts. This would require the coefficient on (for example) Distance to be negative—decreasing the probability of an affirmance—only when the lower court decision was Liberal. The results do not reveal this pattern. For example, consider the results presented in Table 5.4. Among Conservative decisions, those made by proximate lower courts are most likely to be affirmed—but the coefficient on Distance is statistically indistinguishable from 0, revealing no support for the
ideology hypothesis.

The insight that the judicial hierarchy is structured like an organization built for supervision was a leap forward in understanding the American courts. The conception of higher courts as principals who could delegate to their agents in the lower courts, and could then supervise and discipline them as necessary, made great strides. The success of these theories and models in explaining the Supreme Court’s review behavior reinforced the belief that these models were sufficiently comprehensive explanations for Supreme Court justices’ goals. However, the models do not succeed as well when taken to the Supreme Court’s decisions on the merits. As Clarke and Primo (2012) explain, theoretical models’ empirical implications are often not unique—the Supreme Court’s review decisions could be consistent with other, as yet unelucidated theories of judicial decision-making.

There are a few possibilities for the failure to find conclusive evidence at the reversal stage, when such clear evidence tends to exist at the review stage. First, all analyses of Supreme Court decisions on the merits have lower statistical power than studies of certiorari decisions. This is an unavoidable consequence of the Supreme Court’s extremely low rate of review. Second, the Supreme Court’s merits decisions are often relatively inconsequential and more complicated than can be captured in a dichotomous measure (Beim, Cameron and Kornhauser 2012, Harvey and Woodruff 2013). The results on disciplinary hypotheses in particular may be affected by exclusion of cases that were granted, vacated and remanded (as might often be the case in clear instances of non-compliance.) And, of course, measures of Supreme Court and lower court ideology are still somewhat rudimentary and may not always be precise enough to test subtle hypotheses.

Still, the results presented here call out for further investigation. More research should follow the approach of Carrubba et al. (2012): if a model makes predictions about both review and reversal, both should be tested. The results shown here suggest, furthermore,
that there is still ample room for theoretical development in understanding the Supreme Court’s incentives at the dispositional stage. The patterns suggest that the Supreme Court’s dispositions are largely unaffected by ideology of lower court judges. This could be because the Circuit itself is more representative of the content of the rule, and the particular case is unimportant to the Court. Or, patterns may reveal that, as Carrubba et al. (2012) argue, review is for purposes of doctrinal adoption.
Chapter 6

Conclusion

The Supreme Court’s role in the American government is a difficult and complex one. The common law nature of the American judiciary means the Supreme Court is in a position of leading and supervising the disposition of a huge number of disputes each year. As many scholars have acknowledged and studied, this means the Supreme Court must review lower court decisions frequently and wisely to ensure they are complying with doctrine and otherwise resolving cases correctly.

Fewer have considered the implication that the Supreme Court must also frequently and deftly create new law. This is a difficult task riddled with uncertainty, and the Supreme Court’s ability to do it well is critical. In fact, many scholars of comparative jurisprudence have argued that courts’ ability to respond quickly to changing forces leads to economic growth in common law countries. Despite this recognition, however, there has not been comprehensive attention paid to understanding how the Supreme Court crafts law. This dissertation aims to understand how the Supreme Court learns to create law; in particular, it argues the court learns by relying on the Courts of Appeals.

This is a relatively new conception of the hierarchy, in many ways—rather than con-
ceiving of the hierarchy as a top-down control organization, the project conceives of the Supreme Court as learning from its inferiors in a bottom-up informational environment. Many empirical patterns are consistent with the claims of the Supreme Court. For example, the Supreme Court is more likely to review moderate than extreme decisions—even when the extreme decisions are ideologically anathema to it, and made by judges whose ideology would predispose them to decide in that direction. Second, the Supreme Court’s doctrinal choices are affected by the cases it chooses to review—specifically, when lower courts are in conflict, the Supreme Court is more likely to review the doctrines it ultimately strikes down. The Supreme Court’s dispositional votes—its choices of whether to affirm or reverse the decisions of lower courts—are only weakly consistent with the model’s predictions. The Court is slightly more likely to reverse liberal than moderate decisions, but this finding is sensitive to empirical specification and, more importantly, is consistent with many theories of judicial decision-making, not just with a theory of learning.

The theory articulates a very constrained form of learning. In practice, the Supreme Court’s struggle is more complex than the one explored here—justices search for the right sequence of words to achieve the particular doctrinal goals they have. Much of this uncertainty arises because the justices do not know how their opinions will be received, and because they do not know what opinions will achieve the outcomes they seek. Other difficulty arises because, although the justices may have clear expectations for how different terms will be received, there is a high cost to conceiving of new legal language.

For this reason, it is compelling to think that the Supreme Court relies on the Courts of Appeals to bear creative costs. However, existing studies of doctrinal adoption assume the Supreme Court takes good opinions wholesale, while in practice, the Court probably contributes doctrine of its own—it learns something more complex from lower courts’ opinions. Future work in judicial politics should seek to understand more thoroughly what, and how,
the Supreme Court learns. Improvements in natural language processing and automated text analysis may allow for a more nuanced empirical understanding of doctrinal adoption and movement. Theoretical developments in hierarchical communication, persuasion, and learning from inferiors will also prove instructive. Of course, there remains a concern of compliance, and ideological consistency. Supreme Court justices do not seek only to learn, they also seek to achieve goals. A more thorough theory would explicate how they learn to do so.


