Learning in the Judicial Hierarchy

Deborah Beim*
Department of Political Science
Yale University
deborah.beim@yale.edu

July 22, 2013

Abstract

I argue the Supreme Court learns to craft legal rules by relying on the Courts of Appeals as laboratories of law, observing their decisions and reviewing those that best inform legal development. I develop a model that shows how the Supreme Court leverages multiple Courts of Appeals decisions to identify which will be most informative to review, and what decision to make upon review. Because an unbiased judge only makes an extreme decision when there is an imbalance in the parties' evidence, the Supreme Court is able to draw inferences from cases it chooses not to review. Empirically, I then show that, as predicted by the model, the Supreme Court prefers to review moderate decisions rather than extreme ones. 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.

*For generously sharing data, I thank Cliff Carrubba, Tom Clark, John Kastellec, Bill Landes, and Richard Posner.

1 Introduction

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 paper, 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.

I present 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 observe the outcome 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 predicts moderate decisions—where each party prevails on some counts—provide much more information than decisions where one party prevails on all counts. Therefore, if the Supreme Court is attempting to learn, and to develop doctrine, it should be more likely to review these cases. Using a dataset of about 6,000 Courts of Appeals decisions, I show the Supreme Court is more likely to review moderate than extreme decisions. The results contrast much previous literature that argues the Supreme Court audits lower courts and therefore targets extreme decisions. Instead, the results suggest 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 creation of new doctrine—that is, with learning to develop new rules for disposing of future cases.

2 Learning, supervision, and decision making

While the Supreme Court decides under 100 cases per year, the subordinate Courts of Appeals decide tens of thousands of cases each year (United States Courts 2007). 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. Articulating doctrine involves inherent uncertainty (Black and Owens 2012); therefore, understanding the creation of doctrine requires an understanding of how the Supreme Court learns. Several papers have developed theories of judicial learning by considering 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). Those models, like mine, suggest a judge can best articulate doctrine by focusing on the most informative cases.

The Supreme Court's role is not only to articulate doctrine, but to do so from atop a hierarchy. In a hierarchy where multiple agents communicate to principals, agents' messages can interact with—and sometimes counteract—one another, thereby providing more information than the sum of their messages (Epstein 1998, Dewatripont and Tirole 1999, Battaglini 2002, Minozzi 2011). The judicial hierarchy thus 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. The model presented here includes such considerations.

Because the Supreme Court takes so few cases, understanding which Courts of Appeals decisions deserve Supreme Court review, and how to reconcile the inevitable differences that arise between them, is an important question that has received plentiful attention (e.g. Perry 1991, 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 auditing them. A small set of models conceptualizes the judicial hierarchy as a team (Cameron and Kornhauser 2005) trying to resolve cases correctly, but these focus on learning about individual cases, rather than the doctrines that are the primary object of Supreme Court work. With few exceptions (e.g. Cameron 1993, McNollgast 1995, Lax 2003), models of the judicial hierarchy are dyadic—the Supreme Court supervises only one lower court.

A growing body of literature acknowledges that, in reality, the Supreme Court supervises multiple lower courts simultaneously (Lindquist, Haire and Songer 2007) and learns from them. Most Supreme Court opinions cite at least one Courts of Appeals opinion other than the case being reviewed (George and Berger 2005). Repeated experimentation in lower courts is known to aid law creation (Clark and Kastellec 2013), and the Supreme Court allows new legal questions to percolate in the lower courts before resolving them (Klein 2002). Importantly, decisions informing the Supreme Court are often in conflict with one another, which the Supreme Court uses to its advantage. The Rules of the Supreme Court of the United States mention conflict in the lower courts as a reason to consider granting *certiorari*, and indeed, conflict is an excellent predictor of review (Estreicher and Sexton 1984, Caldeira and Wright 1988).

The Supreme Court also seems to adopt doctrine developed in the lower courts. When lower courts are in disagreement, the Supreme Court generally decides in favor of the side that more circuits agree with (Klein and Hume 2003, Lindquist and Klein 2006). Lower courts' citation practices 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 (Corley, Collins and Calvin 2011). Clark and Carrubba (2012) and Carrubba and Clark (2012) argue that because doctrine is costly to produce, the Supreme Court adopts and disseminates rules developed in lower courts.

This paper builds on these findings to understand how they interact. In the model, the Supreme Court aggregates lower court decisions to learn which case to review, then what decision to make upon review. In so doing, the paper speaks to scholarship on strategic communication in hierarchical organizations in general, and to long-standing literatures on the judicial hierarchy in specific.

3 The model

The model consists of a Supreme Court that 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 applicable. 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 motorhome's interior for signs that it functioned as a living space (*California v. Carney*, Justice Stevens, dissenting.) In the model, the two lower courts hear lawyers' arguments for both sides of the dispute, then decide their cases. The Supreme Court observes the decisions the lower courts make, but not the arguments that led to those decisions. Even so, it can draw simple inferences about those arguments from the judges' choices.

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. This allows the Supreme Court to make an informed choice about which case to review, whereupon the Supreme Court will learn what arguments were presented in that case. 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; 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 make, which cases the Supreme Court reviews, and what the Supreme Court does upon review.

3.1 Play of game

The model in this paper uses the architecture from Dewatripont and Tirole (1999). The players are 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. These represent existing doctrines or approaches, which might be thought of as liberal, moderate, and conservative policies, respectively. The Court is 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.[1]

Judges prefer the doctrine that best suits the state of the world, 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, θ_A and θ_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

[1] 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.

sufficient summary of the state is $\theta = \theta_A + \theta_B$. It is common knowledge that:

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

Thus

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

For every state of the world there is an associated doctrine: A if $\theta = -1$, M if $\theta = 0$, and B if $\theta = 1$.

The game proceeds as follows. First, lawyers present evidence to the lower courts about the value of θ. Each lower court then makes a decision based on the evidence he sees. 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 θ and decide whether, and which, case to review. (The Supreme Court can review at most one case.) If the Supreme Court reviews, it learns the arguments that lower court heard, 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.

3.2 Decision making in the lower courts

Simultaneously, the lower courts each hear a case. Both cases depend on the value of θ, which is common across both courts. Because θ cannot be observed directly, this means the judges wish to learn about θ_A and θ_B. Two lawyers—one in each lower court—search for evidence about θ_A.[2] Their searches are independent. The same is true for θ_B: two lawyers,

[2]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.

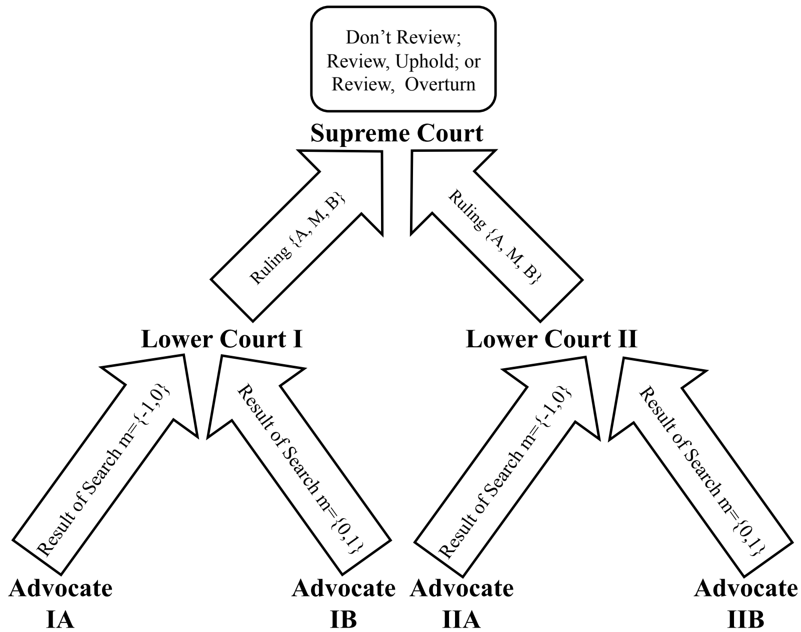

Figure 1: Play of game

one in each lower court, independently search for evidence. Each lawyer then privately presents the results of his search to his judge. m_A denotes the messages of the lawyers for θ_A; m_B denotes the messages of the lawyers for θ_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 $m_i = 0$. As in Che and Kartik (2009), evidence "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 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) (Davis 1942).

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. 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 θ_A presents no hard evidence, this merely 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 judge makes an inference about the value of θ, which incorporates the primitive probability that $|\theta_i| = 1$, α; and the conditional probability that a lawyer's search is successful, q. After establishing a posterior belief about the value of θ, each lower court judge makes a decision, A, M, or B, to correspond to his belief.

3.3 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.[3] The Supreme 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. *Cert* petitions occasionally

[3] In practice, 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 opinion on a particular question.

contain previews of the lawyers' arguments on the merits, but are generally not sufficiently fleshed out for the Supreme Court to determine whether they are valid—this investigation occurs upon review, in reading briefs, hearing oral argument, and deliberating.

After seeing the lower courts' rulings, the Supreme Court updates its beliefs about θ 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 θ. (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 θ, 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. Like the lower court judges, the Supreme Court chooses the doctrine that matches its beliefs about θ. Its decision to reverse or affirm the lower court's ruling follows immediately from this doctrinal choice—it affirms their decision if it agrees based on its own estimate of θ. Of course, the Supreme Court's estimate of θ 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 information provided by the unreviewed lower court's *decision*, which the reviewed lower court cannot see.

3.4 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 θ. Lower court judges update their beliefs based only on their own advocates' messages. Thus, after hearing arguments, LC_I's beliefs about θ are a function of $(\alpha, q, m_{AI}, m_{BI})$ and LC_{II}'s beliefs about θ are a function of $(\alpha, q, m_{AII}, m_{BII})$. The Supreme Court is able to update its beliefs about θ based on both lower courts' decisions. After seeing the lower courts' decisions, the Supreme Court's beliefs about θ are a function of α, q, and the lower courts' decisions. 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. If it reviews LC_I, the Supreme Court's beliefs are a function of $(\alpha, q, m_{AI}, m_{BI})$ and LC_{II}'s decision; if it reviews LC_{II}, its beliefs are a function of $(\alpha, q, m_{AII}, m_{BII})$ and LC_I's decision.

		θ					
		Unbiased			Biased		
		-1	**0**	**1**	**-1**	**0**	**1**
	A	0	-L	-1	0	-L	-1
Doctrine	**M**	-L	0	-L	-L	0	-L
	B	-1	-L	0	-1	-1	0

Table 1: Judges' preferences over doctrine, conditional on the state of the world θ. All judges get 0 from choosing the right doctrine. Mistakes cost -1 or $-L$, where $0 < L < 1$. **Left panel:** Unbiased judges lose more utility from large mistakes than small ones, but have symmetric preferences otherwise. **Right:** For judges biased against B, wrongly choosing B is more costly than wrongly choosing A.

All judges agree on the best doctrine when they know the value of θ with certainty—A when $\theta = -1$, M when $\theta = 0$, and B 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 θ 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 the judges might disagree as to the best outcome if there is uncertainty about whether he is liable: some may believe the manufacturer should not be overburdened with requirements based on inconclusive claims of liability, while others might believe that protecting consumers 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 consumers 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 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 that a biased judge fears a slippery slope, or that a liberal 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. 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.[4]

4 Optimally learning from agents' decisions

As a baseline, I begin by considering 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 4.1. I then consider a scenario where the Supreme Court supervises one ideological ally and one judge who is biased. Section 4.2 presents the equilibrium under these conditions. All proofs appear in the appendix.

4.1 Supervising two unbiased judges

Lower court judges attempt to resolve cases based on the evidence lawyers present. A lower court judge learns the probability of each state, $\theta \in \{-1, 0, 1\}$, from the lawyers'

[4] I 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. See Klein and Hume (2003).

arguments. Recall that before he sees the results of the lawyers' searches for evidence, the judge's prior beliefs are $Pr(\theta_A = -1) = Pr(\theta_B = 1) = \alpha$. Suppose a lawyer's search is unsuccessful, so the judge receives a message $m_i = 0$. Define the judge's posterior belief $pr(\theta_A = -1|m_A = 0) \equiv \hat{\alpha} = \frac{\alpha - \alpha q}{1 - \alpha q}$ (and likewise $pr(\theta_B = 1|m_B = 0) \equiv \hat{\alpha}$). This posterior belief encapsulates the chances that $|\theta_i| = 1$ and the lawyer was simply unsuccessful in proving this.

I restrict $\hat{\alpha} < 1/2$, which ensures that after observing $m_i = 0$, the lower court judge is more inclined to believe that $\theta_i = 0$ than $|\theta_i| = 1$. Then, 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 $\theta = \theta_A + \theta_B$, $\theta_A = -1$, 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.

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 θ. 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 θ 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 sometimes 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 judge 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 upon review the Supreme Court cannot learn enough to want to change the lower courts' decisions. If the courts both decide A or B, the Supreme Court can perfectly infer the messages they received, 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 $\theta = 0$ 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 $\theta_A = -1$ and is slightly more confident that $\theta_B = 0$ 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 $\theta_A = -1$ and less likely that LC_{II} received evidence that $\theta_B = 1$. 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:

$$\begin{cases} A & \textit{iff he receives messages } (-1, 0) \\ B & \textit{iff he receives messages } (0, 1) \\ M & \textit{if he receives messages } (0, 0) \textit{ or } (-1, 1) \end{cases}$$

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

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

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

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

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)*, 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 is 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 moderate decision than the other. 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 choose to review the other case.

As the cost of review rises, though, 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). This is consistent with Black and Owens (2009), who find that the Supreme Court is particularly likely to review extreme conflicts. Furthermore, here the Supreme Court never reviews unless there is conflict.

Finally, even though all judges are identical, the Supreme Court reviews and reverses lower courts' decisions. In fact, if costs are moderate, so that the Supreme Court is willing to review decisions of (A, B) but not when one lower court has decided M, *all* of the Supreme Court's decisions will be reversals, even though the lower courts are perfectly faithful agents.

4.2 Bias in the lower courts

To understand how ideological bias affects learning, consider a scenario in which the Supreme Court is supervising one lower court who shares its unbiased ideological preferences (LC_I) and one lower court who is biased against outcome B (LC_{II}). LC_{II} prefers to choose B if $\theta = 1$, but if $\theta = 0$ he incurs a large loss from choosing B. LC_{II} 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 $(0, 1)$ LC_{II} is not sure enough that $\theta = 1$ to be willing to choose B. This condition is $L < \frac{\hat{\alpha}}{1-\hat{\alpha}}$.[5] Because of this assumption, LC_{II}'s loss from a ruling of B when $\theta = 0$ is larger than that from a ruling of M when $\theta = 1$. Thus, LC_{II} chooses M after seeing $(0, 1)$.

LC_{II}'s bias means the Supreme Court cannot learn as much about θ 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.

[5]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.

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:

$$\begin{cases} A & \textit{iff he receives messages } (-1,0) \\ B & \textit{iff he receives messages } (0,1) \\ M & \textit{if he receives messages } (0,0) \textit{ or } (-1,1) \end{cases}$$

Lower Court II chooses:

$$\begin{cases} A & \textit{iff he receives messages } (-1,0) \\ M & \textit{if he receives messages } (0,1),(0,0) \textit{ or } (-1,1) \end{cases}$$

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 LC_{II} if*
$$c < L[1 - 2\frac{\alpha(1-q)^2}{(1-\alpha)(1-q)^2 + \alpha(1-q)^2 + \alpha q(1-q) + \alpha q^2}],$$
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}'s decision and issues doctrine A.*
  - $(-1,1)$*, 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*
$$c < L\left[1 - 2\frac{\alpha(1-q)^2}{1 - 2q\alpha + 2q^2\alpha}\right],$$
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.*[6] *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.*

[6] See proofs for formula for c^*.

- *If the lower courts chose (B, A) then the Supreme Court takes either case if $c < 2L$, reverses the decision, 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*
$$c < L - 2L\frac{\alpha(1-q)^2}{1-\alpha+\alpha(1-q)(1-q+q^2)},$$
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 homogeneous, unbiased courts in two important ways. First, the probability of review is higher with a biased lower court than without. When one court is biased, the Supreme Court grants *certiorari* in all cases it would review under homogeneity as well as in additional cases. 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 be outcome-consequential. With a biased lower court, however, 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.

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, but the result is more subtle: occasionally the Supreme Court will prefer reviewing its ideological ally to reviewing the biased lower court (such as when the lower courts make decisions (M, A)). This is consistent with recent empirical findings on lower court ideology and Supreme Court review: Lindquist, Haire and Songer (2007) and Walson

(2011) 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, and Clark and Carrubba (2012) show that the Supreme Court prefers to review lower courts that are moderately distant (as opposed to most distant).

Second, the probability of an affirmance is higher when one lower court is biased. Whenever the Supreme Court affirms with two unbiased lower courts, it also affirms when one lower court is biased. It affirms under additional situations because biased decisions are sometimes affirmed in equilibrium. This occurs when the unbiased lower court chooses A, and the biased judge 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_{II}'s lawyers—$(0, 1)$—imply $\theta_B = 1$. Thus, after seeing (A, M) and reviewing LC_{II}'s decision, the Supreme Court knows $\theta = -1 + 1 = 0$. Therefore, even though LC_{II} behaved contrary to how the Supreme Court would have wanted, the Supreme Court upholds his decision.

5 Empirical analysis

5.1 Moderation hypothesis

The model generates a number of empirical predictions. I focus on one: which lower court decision the Supreme Court will choose to review (A, M, or B, representing what one could think of as liberal, moderate, and conservative decisions). 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 to target liberal lower courts' liberal decisions for review and reversal, the learning model predicts that a conservative Supreme Court will review *moderate* decisions. 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 LC_{II}'s decisions of M are most likely to be reviewed (compared to LC_I's decisions and LC_{II}'s other decisions). This is because all other decisions are reviewed conditional on the other court's 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.

5.2 Data

To test this hypothesis, I turn to 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 cases decided by three-judge panels in the Courts of Appeals, including all cases that the Supreme Court reviewed and a stratified random sample of cases that were not reviewed.[7] 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

[7] 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.

as *Conservative* if the conservative party wins (such as the state in a criminal trial, or an employer).[8] When each party prevails on some claims, the decision is coded as *Mixed*.[9] A 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: including as mixed any decision in which the panel voted liberally on one issue and conservatively on another (as Landes and Posner (2009) do), and coding as mixed only those cases that have a mixed outcome on a single dimension.

To test the model's predictions, 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, especially 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 (Cameron, Segal and Songer 2000, Hall 2009, Kastellec 2011*a*). 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 (Bailey 2007). I measure lower court judges' ideology by the party of the president that nominated them. Because the rate of unanimous decisions is so high on the Courts of Appeals, it is best to account for complete panel composition. Therefore, I separately consider each panel type—all Democratic nominees sitting together (abbreviated *DDD*), two Democrats sitting

[8]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. Appendix 2 shows the distribution of decisions across panel types.

[9]For a helpful discussion of when mixed decisions are made, see Lindquist, Haire and Songer (2007).

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. I also present results using Judicial Common Space scores (Epstein et al. 2007) that explicitly place the Courts of Appeals and Supreme Court on the same dimension.

Additionally, I 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). 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).

5.3 Results

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. Only 3.4% of liberal decisions and 1.5% of conservative decisions are reviewed; in contrast, 5.7% of mixed decisions are reviewed. In fact, mixed decisions make up 23% of cases the Supreme Court reviews, even though they make up only 9.8% of cases decided in the Courts of Appeals.

To analyze this further, I estimate a set of logistic regressions of *certiorari* on panel composition and decision direction. Table 2 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.

Model 2 interacts decision direction with panel composition and reveals that, when they make conservative decisions, *DDD* panels are *less* likely to be reviewed than their *RRR* counterparts.[10] 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 subsequent models add relevant control variables: whether there was a *Dissent* in the case, whether the case was about a *Criminal* issue, and *Economic* issue, or neither, the number of amicus briefs, and a time trend. Turning to Model 3, the effect of mixed decisions remains positive and statistically significant and all effects of panel ideology are statistically

[10] This is consistent with Cameron, Segal and Songer (2000), who argue the Supreme Court can be certain it will agree when a court more liberal than it makes a conservative decision.

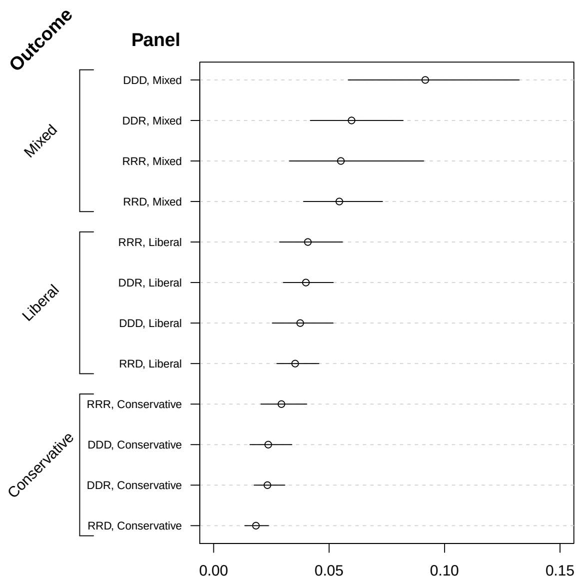

Figure 2: Predicted probability of review by panel type and decision. Control variables are held at their means. Mixed decisions are most likely to be reviewed, especially those made by non-allies.

indistinguishable from 0. Most importantly, the total effect of mixed decisions by *DDD* panels, 1.4, is significantly larger than that of liberal decisions by *DDD* panels, .43.

Figure 2 translates these effects into predicted probabilities of *certiorari*, by decision direction and panel composition. The figure makes it clear that decision is a more influential variable than panel composition. Mixed decisions are by far the most likely cases to be reviewed, regardless of the panel that made them. 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. *DDD* panels' *liberal* decisions have a 3% chance of review; while their *mixed* decisions have an 8% probability of review. This is a sizable increase; the difference between two rates is distinguishable from 0 at $p < .05$. The high rate of review of *DDD* panels' mixed decisions is particularly noteworthy given that *certiorari* is so rarely granted—the overall probability is a mere 2.5%.

Models 4 and 5 show the results are robust to alternative specifications of lower court ideology. Model 4 replaces panel composition with the median Judicial Common Space score of the judges on the panel (Giles, Hettinger and Peppers 2001; 2002, Epstein et al. 2007). Since some Democratic appointees are more liberal than others (and likewise for Republican nominees), these scores account for ideological heterogeneity within party.[11] Again, mixed decisions are more likely to be reviewed than liberal decisions. Also, liberal panels—even when they make liberal decisions—are no more likely to be reviewed than conservative panels. Model 5 excludes the Southern circuits—the 5th and 11th—where presidential party is likely to perform worst, since some Democratic-appointed judges are likely more conservative than their Republican counterparts in the North. This exclusion strengthens the positive effect of effect of mixed decisions and renders the interaction between mixed decisions and *DDD* panels significant.

Model 6 restricts the *Mixed* coding to those cases which were found to be mixed on a single dimension. The effect persists—mixed decisions are still most likely to be reviewed. This suggests that the effect of mixed decisions is not attributable only to their being more complex than liberal and conservative decisions. Decisions that present only one question but reach a mixed outcome on it are more likely to be reviewed than purely liberal or conservative decisions.

[11]Specifically, Judicial Common Space scores assign the NOMINATE score of the judge's home-state senator, when that senator is of the same party as the president, and the president's NOMINATE Common Space score when he is not.

	(1)	(2)	(3)	(4)	(5)	(6)
	Two Issues	Two Issues	Two Issues	Two Issues	Non-south only	Primary Issue
Mixed	**1.36***	**0.93***	**0.77***	**1.00***	**0.79***	**0.80***
	(0.06)	(0.27)	(0.34)	(0.13)	(0.33)	(0.37)
Liberal	**0.82***	**0.49***	0.50	**0.51***	**0.53***	0.51
	(0.08)	(0.20)	(0.27)	(0.11)	(0.27)	(0.26)
DDD		**-0.42***	-0.08		-0.10	0.01
		(0.21)	(0.29)		(0.30)	(0.28)
DDR		**-0.40***	-0.08		-0.06	-0.03
		(0.16)	(0.25)		(0.25)	(0.24)
RRD		**-0.59***	-0.32		-0.27	-0.24
		(0.17)	(0.25)		(0.24)	(0.24)
Panel Median				-0.15		
				(0.26)		
DDD, Mixed		**0.80***	0.71		**0.81***	0.55
		(0.37)	(0.42)		(0.43)	(0.45)
DDR, Mixed		0.39	0.27		0.32	0.21
		(0.31)	(0.35)		(0.35)	(0.38)
RRD, Mixed		0.53	0.40		0.39	0.30
		(0.31)	(0.36)		(0.36)	(0.41)
DDD, Liberal		0.30	0.01		0.11	-0.07
		(0.29)	(0.34)		(0.36)	(0.33)
DDR, Liberal		0.37	0.11		0.16	0.06
		(0.23)	(0.27)		(0.28)	(0.26)
RRD, Liberal		**0.47***	0.23		0.22	0.17
		(0.24)	(0.29)		(0.29)	(0.28)
Panel Median × Mixed			-0.10			
			(0.45)			
Panel Median × Liberal			-0.03			
			(0.34)			
Dissent			0.67	**1.36***	0.59	0.68
			(0.48)	(0.11)	(0.46)	(0.49)
Time			-0.01	**-0.02***	-0.01	-0.01
			(0.01)	(0.01)	(0.01)	(0.01)
Criminal			**-0.55***	**-0.43***	**-0.54***	**-0.56***
			(0.11)	(0.12)	(0.11)	(0.11)
Economic			**-0.25***	-0.20	**-0.31***	**-0.24***
			(0.09)	(0.10)	(0.09)	(0.09)
Number Amicus Briefs			**0.33***	**0.37***	**0.32***	**0.34***
			(0.05)	(0.05)	(0.05)	(0.05)
Intercept	**-4.14***	**-3.74***	**-3.68***	**-3.67***	**-3.79***	**-3.70***
	(0.06)	(0.13)	(0.37)	(0.32)	(0.37)	(0.36)
Circuit Fixed Effects	No	No	Yes	No	Yes	Yes
N	6,289	6,289	6,289	4,543	5,689	6,289

Table 2: 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.

The results contrast existing disciplinary theories of the judicial hierarchy. Previous literature on the judicial hierarchy argues that conservative courts wish to minimize liberal *decisions* in the lower courts, and will therefore review liberal courts' liberal decisions. Empirically, however, the conservative Burger Court targeted mixed rather than purely liberal decisions. The results suggest that the Supreme Court wishes to minimize liberal *lawmaking* in the lower courts, as opposed to liberal decisions. The Supreme Court is most likely to review the decisions that were likely to have been difficult and will therefore offer learning opportunities.

6 Discussion and Conclusion

Law is not static. As society changes, new problems emerge, new classes of disputes arise, and doctrine must adapt to govern their resolution. This paper 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, Klein and Hume 2003, Baker and Mezzetti 2012, Corley, Collins and Calvin 2011, Niblett 2013, Carrubba et al. 2012, Clark and Kastellec 2013). This approach contrasts with the dominant mode 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 encourage division of labor: the Supreme Court may specialize in answering difficult questions while relying on agents to answer easier ones (Garicano and Zandt 2013). This model, however, conceives of the Supreme Court as specializing in *supervision* rather than in the resolution of similar, but more difficult, questions.

The paper 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 doctrinal 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 does learn from lower courts' decisions. In addition to the review predictions tested in this paper, future work might test the model's implications for the Supreme Court's doctrinal decisions and dispositional choices. For example, the model predicts mixed decisions are less likely to be reversed than either liberal or conservative decisions, which could easily be tested.

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 for information and write a dissenting opinion that presents the evidence he finds.[12] Similarly, in the model presented here an affirmance is doctrinally useful, whereas in disciplinary models, they occur only as an accident of incomplete information. This alternative intuition might better explain the Supreme Court's opinions affirming decisions of lower courts. Affirmances are mistakes in disciplinary models, so they can be assumed to yield short opinions without much argumentation; here they are equally effective vehicles for communicating doctrine.

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 paper contributes to a broader literature on information-gathering and optimal experimentation in hierarchical organizations. Two 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? Here, 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, this model ends once the Supreme Court chooses a doctrine. In practice,

[12] 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.

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.

References

Bailey, Michael A. 2007. "Comparable Preference Estimates across Time and Insitutions for the Court, Congress and Presidency." *American Journal of Political Science* 51(3):433–48.

Baker, Scott and Claudio Mezzetti. 2012. "A Rational Theory of Jurisprudence." *Journal of Political Economy* 120(3):513 – 551.

Battaglini, Marco. 2002. "Multiple Referrals and Multidimensional Cheap Talk." *Econometrica* 70(4):1379–1401.

Beim, Deborah, Alexander V. Hirsch and Jonathan P. Kastellec. 2012. "Whistleblowing and Compliance in the Judicial Hierarchy." Princeton University working paper.

Black, Ryan C. and Ryan J. Owens. 2009. "Agenda Setting in the Supreme Court: The Collision of Policy and Jurisprudence." *Journal of Politics* 71(3):1062–1075.

Black, Ryan C. and Ryan J. Owens. 2012. "Looking Back to Move Forward: Quantifying Policy Predictions in Political Decision Making." *American Journal of Political Science* 56(4):802–816.

Caldeira, Greg A., John R. Wright and Christopher J. W. Zorn. 1999. "Sophisticated Voting and Gate-Keeping in the Supreme Court." *Journal of Law Economics & Organization* 15:549–72.

Caldeira, Gregory A. and John R. Wright. 1988. "Organized Interests and Agenda Setting in the U.S. Supreme Court." *The American Political Science Review* 82(4):1109–1127.

Calvert, Randall L. 1985. "The Value of Biased Information: A Rational Choice Model of Political Advice." *Journal of Politics* 47(2):530–555.

Cameron, Charles and Lewis Kornhauser. 2005. Appeals Mechanisms, Litigant Selection and the Structure of Judicial Hierarchies. In *Institutional Games and the Supreme Court*, ed. Jon Bond, Roy Fleming and James Rogers. University of Virginia Press.

Cameron, Charles M. 1993. "New Avenues for Modeling Judicial Politics." Paper presented at the Conference on the Political Economy of Public Law, Rochester, N.Y.

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

Carnes, Nicholas. 2007. "logit.survey: Survey-Weighted Logistic Regression for Dichotomous Dependent Variables in Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," `http://gking.harvard.edu/zelig`.".

Carrubba, Cliff, Barry Friedman, Andrew D. Martin and Georg Vanberg. 2012. "Who Controls the Content of Supreme Court Opinions?" *American Journal of Political Science* 56(2):400–412.

Carrubba, Clifford J. and Tom S. Clark. 2012. "Rule Creation in a Political Hierarchy." *American Political Science Review* 106(3):622–643.

Che, Yeon-Koo and Navin Kartik. 2009. "Opinions as Incentives." *Journal of Political Economy* 117(5).

Clark, Tom S. 2009. "A Principal Agent Theory of En Banc Review." *Journal of Law, Economics & Organization* 25(1):55–79.

Clark, Tom S. and Clifford J. Carrubba. 2012. "A Theory of Opinion Writing in the Judicial Hierarchy." *Journal of Politics* 74(2):584–603.

Clark, Tom S. and Jonathan P. Kastellec. 2013. "The Supreme Court and Percolation in the Lower Courts: An Optimal Stopping Model." *Journal of Politics* 75(1):150–168.

Cooter, Robert, Lewis Kornhauser and David Lane. 1979. "Liability Rules, Limited Information, and the Role of Precedent." *The Bell Journal of Economics* 10(1):366–373.

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

Davis, Kenneth C. 1942. "An Approach to Problems of Evidence in the Administrative Process." *Harvard Law Review* 55(364).

Dewatripont, Mathias and Jean Tirole. 1999. "Advocates." *Journal of Political Economy* 107(1):1–39.

Epstein, David. 1998. "Partisan and Bipartisan Signaling in Congress." *Journal of Law Economics & Organization* 14(2).

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

Epstein, Lee, William M. Landes and Richard A. Posner. 2011. "Why (and When) Judges Dissent: A Theoretical and Empirical Analysis." *Journal of Legal Analysis* 3(1):101–137.

Estreicher, Samuel and John E. Sexton. 1984. "A Managerial Theory of the Supreme Court's Responsibilities: An Empirical Study." *New York University Law Review* 59(4).

Friedman, Barry. 2006. "Taking Law Seriously." *Perspectives on Politics* 4(2):261–76.

Gailmard, Sean and John Patty. 2013. "Participation, Process, & Policy: The Informational Value of Politicized Judicial Review." Washington University in St. Louis Working Paper.

Garicano, Luis and Timothy Van Zandt. 2013. Hierarchies and the Division of Labor. In *The Handbook of Organizational Economics*, ed. Robert Gibbons and John Roberts. Princeton University Press.

George, Tracey E. and Jeffrey A. Berger. 2005. "Judicial Entrepreneurs on the U.S. Courts of Appeals: A Citation Analysis of Judicial Influence." Available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=789544`.

Giles, Micheal W., Virginia A. Hettinger and Todd Peppers. 2001. "Picking Federal Judges: A Note on Policy and Partisan Selection Agendas." *Political Research Quarterly* 54:623–41.

Giles, Micheal W., Virginia A. Hettinger and Todd Peppers. 2002. "Measuring the Preferences of Federal Judges: Alternatives to Party of the Appointing President." Emory University working paper.

Gryski, Gerald S. and Gary Zuk. 2008. "A Multi-User Data Base on the Attributes of U.S. Appeals Court Judges, 1801-2000." Available at http://www.cas.sc.edu/poli/juri/auburndata.htm, accessed 15 August 2008.

Gryski, Gerald S., Gary Zuk and Sheldon Goldman. 2008. "A Multi-User Data Base on the Attributes of U.S. District Court Judges, 1789-2000." Available at http://www.cas.sc.edu/poli/juri/auburndata.htm, accessed 15 August 2008.

Hall, Matthew. 2009. "Experimental Justice: Random Judicial Assignment and the Partisan Process of Supreme Court Review." *American Politics Research* 37(2):195–226.

Hansford, Thomas G., James F. Spriggs and Anthony Stenger. 2010. "The Information Dynamics of Vertical Stare Decisis." Available at `http://ssrn.com/abstract=1580528`.

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

Kastellec, Jonathan P. 2007. "Panel Composition and Judicial Compliance on the U.S. Courts of Appeals." *Journal of Law, Economics & Organization* 23(2):421–441.

Kastellec, Jonathan P. 2011*a*. "Hierarchical and Collegial Politics on the U.S. Courts of Appeals." *Journal of Politics* 73(2):345–61.

Kastellec, Jonathan P. 2011*b*. "Panel Composition and Voting on the U.S. Courts of Appeals Over Time." *Political Research Quarterly* 64(2):377–91.

King, Gary and Langche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 6(1):137–63.

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

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

Kornhauser, Lewis A. 1989. "An Economic Perspective on Stare Decisis." *Chicago-Kent Law Review* 65(63).

Landes, William M. and Richard A. Posner. 2009. "Rational Judicial Behavior: A Statistical Study." *Journal of Legal Analysis* 1(2).

Lax, Jeffrey R. 2003. "Certiorari and Compliance in the Judicial Hierarchy: Discretion, Reputation and the Rule of Four." *Journal of Theoretical Politics* 15:61–86.

Lindquist, Stefanie A. and David E. Klein. 2006. "The Influence of Jurisprudential Considerations on Supreme Court Decisionmaking: A Study of Conflict Cases." *Law and Society Review* 40(1).

Lindquist, Stefanie A., Susan B. Haire and Donald R. Songer. 2007. "Supreme Court Auditing of the US Courts of Appeals: An Organizational Perspective." *Journal of Public Administration Research and Theory* 17(4):607–624.

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

McNollgast. 1995. "Politics and the Courts: A Positive Theory of Judicial Doctrine and the Rule of Law." *Southern California Law Review* 68:1631–89.

Minozzi, William. 2011. "A Jamming Theory of Politics." *Journal of Politics* 73(2):301–315.

Niblett, Anthony. 2013. "Case-by-Case Adjudication and the Path of the Law." *Journal of Legal Studies* 42.

Perry, Jr., H.W. 1991. *Deciding to Decide: Agenda Setting in the United States Supreme Court*. Cambridge: Harvard University Press.

Shapiro, Carolyn. 2006. "The Limits of the Olympian Court: Common Law Judging versus Error Correction in the Supreme Court." *Washington and Lee Law Review* 63:271.

Songer, Donald R. 1999. "The United States Courts of Appeals Database, 1925-1996." available at `http://www.as.uky.edu/academics/departments_programs/PoliticalScience/PoliticalScience/FacultyResources/Resources/Ulmer/Pages/USCourtsofAppealsDatabase.aspx`, accessed 3 June 2008.

Songer, Donald R. 2008. "Phase II Courts of Appeals Database." Available at http://www.cas.sc.edu/poli/juri/appct.htm; Version updated October 2008.

Spaeth, Harold J. 2011. "Supreme Court Database." http://scdb.wustl.edu/data.php.

Spitzer, Matthew and Eric Talley. 2012. "Left, Right, and Center: Strategic Information Acquisition and Diversity in Judicial Panels." *Journal of Law Economics & Organization* Forthcoming.

Tanenhaus, Joseph, Marvin Schick, Matthew Muraskin and Daniel Rosen. 1963. The Supreme Court's Certiorari Decisions: Cue Theory. In *Judicial Decision Making*, ed. Glendon Schubert. Glencoe, IL: Free Press pp. 111–32.

United States Courts, Administrative Office of the. 2007. "Federal Judicial Caseload Statistics, March 31, 2007." Available at `http://www.uscourts.gov/caseload2007/contents.html`.

Walson, John F. 2011. "Learning by Review." Emory University working paper.

Xie, Yu and Charles F. Manski. 1989. "The Logit Model and Response-Based Samples." *Sociological Methods and Research* 17(3):283–302.

Appendix I: 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 $pr(\theta = -1) = pr(\theta = 1) = \frac{(1-q)\alpha}{1+\alpha-2q\alpha}$. He believes $pr(\theta = 0) = \frac{1-\alpha}{1+\alpha-2q\alpha}$. Therefore, his expected utility from choosing A or B is

$-\frac{(1-q)\alpha}{1+\alpha-2q\alpha} - L\frac{1-\alpha}{1+\alpha-2q\alpha}$. His expected utility from choosing M is $-L\frac{\alpha(1-q)}{1+\alpha-2q\alpha} - L\frac{\alpha(1-q)}{1+\alpha-2q\alpha}$. His expected utility from M is therefore greater:

$$\begin{aligned} & -L\tfrac{\alpha(1-q)}{1+\alpha-2q\alpha} - L\tfrac{\alpha(1-q)}{1+\alpha-2q\alpha} - \left[-\tfrac{(1-q)\alpha}{1+\alpha-2q\alpha} - L\tfrac{1-\alpha}{1+\alpha-2q\alpha}\right] \\ = \quad & -L\alpha(1-q) + \alpha(1-q) - L\alpha(1-q) + L(1-\alpha) \\ = \quad & L(1-\alpha) + (1-q)(\alpha - 2L\alpha) > 0. \end{aligned}$$

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 3 shows all possible histories and the utility change from review.

State of the world	Messages	Utility change
$\theta = -1$	(0,0;0,0)	-2L without review, -2L with review
$\theta = -1$	(-1,1;0,0)	Cannot happen
$\theta = -1$	(-1,1;-1,1)	Cannot happen
$\theta = 1$	(0,0;0,0)	-2L without review, -2L with review
$\theta = 1$	(-1,1;0,0)	Cannot happen
$\theta = 1$	(-1,1;-1,1)	Cannot happen
$\theta = 0 + 0$	(0,0;0,0)	0 without review, 0 with review
$\theta = 0 + 0$	(-1,1;0,0)	Cannot happen
$\theta = 0 + 0$	(-1,1;-1,1)	Cannot happen
$\theta = -1 + 1$	(0,0;0,0)	0 without review, 0 with review
$\theta = -1 + 1$	(-1,1;0,0)	0 without review, 0 with review
$\theta = -1 + 1$	(-1,1;-1,1)	0 without review, 0 with review

Table 3: 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:

State of the world	Messages	Utility change
$\theta = -1$	(-1,0;0,0)	-L without review, 0 with review
$\theta = -1$	(-1,0;-1,1)	Cannot happen
$\theta = -1 + 1$	(-1,0;0,0)	-L without review, -2L with review
$\theta = -1 + 1$	(-1,0;-1,1)	-L without review, 0 with review

Thus the Supreme Court's utility without review is $-L$. The Supreme Court's utility if it reviews is $-c - 2L * Pr(\theta = -1 + 1; -1, 0; 0, 0|A, M)$. Therefore the Supreme Court will review if

$$
\begin{aligned}
&-c - 2L\tfrac{\alpha^2(1-q)^3 q}{Pr(A,M)} > -L \\
&L - 2L\tfrac{\alpha^2(1-q)^3 q}{Pr(A,M)} > c \\
&c < L\left[1 - 2\tfrac{\alpha^2(1-q)^3 q}{Pr(A,M)}\right] \\
&c < L\left[1 - 2\tfrac{\alpha^2(1-q)^3 q}{\alpha(1-\alpha)q(1-q)+\alpha^2 q(1-q)^3+\alpha^2 q^3(1-q)}\right] \\
&c < L\left[1 - 2\tfrac{\alpha(1-q)^2}{1-2q\alpha+2q^2\alpha}\right]
\end{aligned}
$$

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:

$$EU_{LC_{II}}[B] = -1 * Pr(\theta = 0|0,1) = -\hat{\alpha}$$

$$EU_{LC_{II}}[M] = -L * Pr(\theta = 1|0,1) = -L(1-\hat{\alpha})$$

$$EU_{LC_{II}}[A] = -1$$

So LC_{II} chooses M so long as $L < \frac{\hat{\alpha}}{1-\hat{\alpha}}$, which I assume henceforth.[13] 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.[14]

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

[13]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.

[14]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:

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

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 $\theta = 0$.

State of the world	Messages	Utility change
$\theta = -1$	(-1,0;0,0)	-L without review, 0 with review
$\theta = -1$	(-1,0;0,1)	Cannot happen
$\theta = -1$	(-1,0;-1,1)	Cannot happen
$\theta = -1 + 1$	(-1,0;0,0)	-L without review, -2L with review
$\theta = -1 + 1$	(-1,0;0,1)	-L without review, 0 with review
$\theta = -1 + 1$	(-1,0;-1,1)	-L without review, 0 with review

Therefore its utility from review is $-c - 2LPr(\theta = -1 + 1; -1, 0; 0, 0|B, M)$.

It will review so long as

$$-c - 2LPr(\theta = -1 + 1; -1, 0; 0, 0|B, M) > -L$$

$$L - 2LPr(\theta = -1 + 1; -1, 0; 0, 0|B, M) > c$$

$$c < L - 2LPr(\theta = -1 + 1; -1, 0; 0, 0|B, M)$$

$$c < L[1 - 2\frac{Pr(\theta=-1+1;-1,0;0,0)}{Pr(B,M)}]$$

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

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 $\theta_A = -1$.

Its utility gain from reviewing is as follows:

State of the world	Messages	Utility change
$\theta = 1$	(0,1;0,0)	-L without review, 0 with review
$\theta = 1$	(0,1;0,1)	-L without review, 0 with review
$\theta = 1$	(0,1;-1,1)	Cannot happen
$\theta = -1 + 1$	(0,1;0,0)	-L without review, -2L with review
$\theta = -1 + 1$	(0,1;0,1)	-L without review, -2L with review
$\theta = -1 + 1$	(0,1;-1,1)	-L without review, 0 with review

Therefore it will review if:

$$-L < -c - 2L * Pr(\theta = -1 + 1; 0, 1; 0, \dot{|}B, M)$$

$$-L < -c - 2L\frac{\alpha^2(1-q)^2 q}{Pr(B,M)}$$

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

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

After (M, M): The Supreme Court puts positive probability on all values of θ. The utility change from reviewing is as follows:

State of the world	Messages	Utility change
$\theta = -1$	(0,0;0,0)	-2L without review, -2L with review
$\theta = -1$	(0,0;0,1)	Cannot happen
$\theta = -1$	(0,0;-1,1)	Cannot happen
$\theta = -1$	(-1,1;0,0)	Cannot happen
$\theta = -1$	(-1,1;0,1)	Cannot happen
$\theta = -1$	(-1,1;-1,1)	Cannot happen
$\theta = 1$	(0,0;0,0)	-2L without review, -2L with review
$\theta = 1$	(0,0;0,1)	-2L without review, 0 with review
$\theta = 1$	(0,0;-1,1)	Cannot happen
$\theta = 1$	(-1,1;0,0)	Cannot happen
$\theta = 1$	(-1,1;0,1)	Cannot happen
$\theta = 1$	(-1,1;-1,1)	Cannot happen
$\theta = -1 + 1$	(0,0;0,0)	0 without review, 0 with review
$\theta = -1 + 1$	(0,0;0,1)	0 without review, -2L with review
$\theta = -1 + 1$	(0,0;-1,1)	0 without review, 0 with review
$\theta = -1 + 1$	(-1,1;0,0)	0 without review, 0 with review
$\theta = -1 + 1$	(-1,1;0,1)	0 without review, -2L with review
$\theta = -1 + 1$	(-1,1;-1,1)	0 without review, 0 with review
$\theta = 0 + 0$	(0,0;0,0)	0 without review, 0 with review
$\theta = 0 + 0$	(0,0;0,1)	Cannot happen
$\theta = 0 + 0$	(0,0;-1,1)	Cannot happen
$\theta = 0 + 0$	(-1,1;0,0)	Cannot happen
$\theta = 0 + 0$	(-1,1;0,1)	Cannot happen
$\theta = 0 + 0$	(-1,1;-1,1)	Cannot happen

Therefore, the Supreme Court's expected utility without review is

$$-2L * Pr(\theta = 1; 0, 0; 0, 1|M, M) = -2L\frac{\alpha(1-\alpha)q(1-q)}{Pr(M,M)}$$

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

The Supreme Court's expected utility with review is

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

The Supreme Court will review if:

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

$$>$$

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

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

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

Otherwise it will not review. ■

Appendix 2: 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 (2011*b*). 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 4 presents the distribution of panel composition by case outcome.

	Unspecifiable	Conservative	Mixed	Liberal	Total
DDD	.01	.05	.02	.05	898
DDR	.03	.17	.05	.15	2740
RRD	.03	.16	.04	.12	2432
RRR	.01	.06	.01	.04	901
Total	587	3054	834	2496	6971

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