

# Learning in the Judicial Hierarchy

Deborah Beim\*

Department of Political Science

Yale University

deborah.beim@yale.edu

July 5, 2016

## 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. 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.

---

\*I thank Chuck Cameron, Brandice Canes-Wrone, Tom Clark, Justin Fox, John Kestellec, Lewis Kornhauser, Adam Meirowitz, Kelly Rader, Jim Rogers, Eric Dickson, anonymous reviewers, and seminar participants at Emory University, New York University, Princeton University and the 2012 conference of the Midwest Political Science Association for helpful comments. Shikhar Singh provided invaluable research assistance.

# 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 simultaneously helps to explain stylized facts about review and to illuminate new

areas for empirical study. For example, it has long been known that the Supreme Court reviews in order to resolve conflicts in the lower courts, which is true in the model presented here. But the model here shows how not all cases are equally good vehicles for resolving lower court conflict. Instead, the Supreme Court will be more likely to review cases from the side of the conflict it ultimately disagrees with—a result that is consistent with the patterns presented in Wasby (2005) and Summers and Newman (2011). Similarly, it has long been known that although the Supreme Court is more likely to review the decisions of judges who are ideologically distant, it also frequently reviews and reverses the decisions of its ideological allies (Walson 2011). The model presented here offers a rationalization for these phenomena. It also makes a series of novel empirical predictions, including *which* case will be most informative and what the probability of reversal will be conditional on review.

The model predicts decisions in which each party prevails on some counts are more informative to review than decisions where one party prevails on all counts. Therefore, the Supreme Court should be more likely to review these cases. The results contrast much previous literature that argues the Supreme Court audits lower courts and therefore targets decisions likely to be wrong. Instead, the results suggest the Court reviews cases that offer learning opportunities, since it is not uniquely concerned with ensuring that lower courts follow established doctrine—that is, with deciding easy cases as the Supreme Court would like. The Court also reviews to create new doctrine and 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 struc-

turing 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 2014). 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.

Most models of the judicial hierarchy are dyadic—the Supreme Court supervises only one lower court (though see Cameron 1993, McNollgast 1995, Lax 2003). 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 Kastlelec 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 Supreme Court Rules 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 lower courts' decisions, 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 doc-

trine. 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.

### 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.<sup>1</sup>

Judges prefer the doctrine that best suits the state of the world. Because the area of law is relatively new, they do not know which doctrine 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:

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

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

---

<sup>1</sup> 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.

$B$  if  $\theta = 1$ , and:

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

The game proceeds as follows. First, lawyers present evidence to the lower courts about the value of  $\theta$ . 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  $\theta$  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  $\theta$ , which is common across both courts. Because  $\theta$  cannot be observed directly, this means the judges wish to learn about  $\theta_A$  and  $\theta_B$ . Two lawyers—one in each lower court—search for evidence about  $\theta_A$ .<sup>2</sup> 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 judge.  $m_A$  denotes the messages of the lawyers for  $\theta_A$ ;  $m_B$  denotes the messages of the

---

<sup>2</sup> I discuss the game as if lawyers are presenting evidence to the court, but treat them mechanically rather than analyzing their behavior. The behavior of advocates is 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.

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  $m_i = 0$ . Evidence is dispositive about the state of the world, not just the case in question. Legally, these 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  $\theta_A$  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  $\theta$ , 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 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.<sup>3</sup> 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 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  $\theta$  and decides whether to review either of the lower courts' decisions. If the Supreme Court chooses not to review, the lower courts' decisions stand and the game ends. The Supreme Court cannot rule on an issue without reviewing at least one case. 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, hearing arguments, and writing an opinion. 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

---

<sup>3</sup> In practice, the Supreme Court may consolidate cases and hear them together. I consider this possibility in Proposition 3.

the decision not reviewed. 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 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 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  $\theta$ . A lower court judge updates his beliefs based only on the messages sent by his own advocates. Thus, after hearing arguments,  $LC_I$ 's beliefs about  $\theta$  are a function of  $(\alpha, q, m_{AI}, m_{BI})$  and  $LC_{II}$ 's beliefs about  $\theta$  are a function of  $(\alpha, q, m_{AII}, m_{BII})$ . 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  $\alpha, q$ , 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.

All judges agree on the best doctrine when they know the value of  $\theta$  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  $\theta$  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 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.<sup>4</sup>

## 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’ arguments.

---

<sup>4</sup> 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).

Recall that before he sees the results of the lawyers' searches for evidence, the judge's prior beliefs are  $Pr(\theta_A = -1) = Pr(\theta_B = 1) = \alpha$ .

I restrict  $q > 2 - 1/\alpha$ , 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  $q > 2 - 1/\alpha$  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  $\theta$ . The lower court judge believes it is more likely that  $\theta = 0$  than either  $-1$  or  $1$ . Furthermore, 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 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 drives 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  $\theta$ —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, as hearing the case and writing the opinion will still be time consuming. 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:

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

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

- If the lower courts chose  $(A, A)$ ,  $(B, B)$ , or  $(M, M)$ , the Supreme Court does not review a case.
- If the lower courts chose  $(A, M)$ , the Supreme Court pays  $c$  to review  $LC_{II}$  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 pays  $c$ , reviews a case (either case), reverses the decision, and issues universal precedent  $M$ .

The most notable result here 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 neither too high nor too low, 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 Supreme Court 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  $L$  so that after seeing  $(0, 1)$   $LC_{II}$  is not sure enough that  $\theta = 1$  to be willing to choose  $B$ . This condition is  $L < \frac{\alpha(1-q)}{1-\alpha}$ .<sup>5</sup> 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  $\theta$  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:*

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

*Lower Court II chooses:*

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

---

<sup>5</sup> 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.

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 pays  $c$  to review  $LC_{II}$  if
 
$$c < L \left[ 1 - 2 \frac{\alpha^2 q (1 - q)^3}{\alpha(1 - \alpha)q(1 - q) + \alpha^2 q (1 - q)^3 + \alpha^2 q^2 (1 - q)^2 + \alpha^2 q^3 (1 - q)} \right],$$
 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 pays  $c$  to review  $LC_I$  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 pays  $c$  to review  $LC_{II}$  if  $\alpha < 2q(1 - q)$  and  $c < c^*(L, \alpha, q)$ , otherwise it does not review a case.<sup>6</sup> 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 decision, and issues doctrine  $M$ . If  $c \geq 2L$ , it does not review.
- If the lower courts chose  $(B, M)$  then the Supreme Court pays  $c$  to review  $LC_{II}$  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  $M$ .

---

<sup>6</sup> See proofs for formula for  $c^*$ .

–  $(0, 1)$ , 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 empirical findings on lower court ideology and Supreme Court review: Lindquist, Haire and Songer (2007) shows 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.

This ability to leverage information contained in a decision that is never reviewed has important implications for how many cases the Supreme Court has to hear. 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,<sup>7</sup> 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 almost 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.)

**Proposition 3 (Case Consolidation)** *With homogeneous lower courts, the Supreme Court will never choose to consolidate cases. With heterogeneous lower courts, the Supreme Court only chooses to consolidate cases if both have made a decision of  $M$ .*

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

---

<sup>7</sup>According 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.

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. In other words, there is rarely a need for the Supreme Court to look at the arguments and evidence from all cases, even if lower courts are assumed to learn nothing from one another.

Proposition 3 shows that the Supreme Court almost never chooses to consolidate cases, even when it is given the opportunity to do so. Only when heterogeneous lower courts both make decisions of  $M$  might the Supreme Court choose to review both cases. In other words, there is rarely a need for the Supreme Court to look at the arguments and evidence from all cases, even if lower courts are assumed to learn nothing from one another.

### **4.3 Learning, good law, and lower court diversity**

Much research on learning from subordinates suggest that increased diversity improves the informational environment. Preference diversity allows a principal to strategically delegate tasks to the best-suited agent (Battaglini 2002), decreases the opportunity to lie (Epstein 1998, Beim, Hirsch and Kastellec 2014), decreases free-riding, and decreases cascades (Austen-Smith and Banks 1996, Daughety and Reinganum 1999). This model, however, illustrates a negative consequence of diversity: some diversity is actually bias, and bias can hamper information transmission by silencing signals. When lower courts become biased and therefore communicate less information to their superiors, they are much less useful and the principal's utility declines. This result contrasts Callander and Harstad (2015), who find that heterogeneity increases the incentive to experiment, thereby providing more (if overly extreme) learning opportunities.

**Remark 4** *Supervising a biased lower court causes the Supreme Court to conduct useless review.*

Here, diversity increases the Supreme Court’s workload without offering any informational advantages. Diversity has two negative consequences for the Supreme Court. First, as described in the Remark above, the Supreme Court must review more often to compensate for the biased lower court obscuring information. Most notably, when both lower courts receive balanced signals (either  $0, 0$  or  $-1, 1$ ) the Supreme Court should not review and, when lower courts are unbiased, does not review. But under these circumstances the Supreme Court *does* review when one lower court is biased, a cost it incurs to check that the biased court has not made a biased decision. This can be interpreted as a waste of the Supreme Court’s time. Furthermore, because of the obfuscation, the Supreme Court might miss the opportunity to issue an obvious  $M$  decision after the lower courts receive messages  $(-1, 0; 0, 1)$  (that is, after unbiased lower courts would issue decisions  $(A, B)$ ).

When

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

it is too costly to check whether the biased lower court saw  $(0, 1)$  so the Supreme Court misses the opportunity to enforce the right decision.

#### 4.4 Learning within lower courts

While the previous results assume lower courts cannot observe one another’s decisions, in practice the Courts of Appeals may learn from each other. Under what conditions does a peer circuit’s decision affect a lower court judge’s choice? How does this affect the Supreme Court’s informational environment?

When lower courts learn from one another, the probability of review falls considerably. The need for learning-based review falls even further. The reason is the following: the bulk of the Supreme Court’s work is supervisory. When lower courts can learn from one another, a large swath of the Supreme Court’s workload is taken over by appellate court judges. Consider, for example, a lower court who observes messages  $(0, 1)$  and knows a previous lower court has made

a decision of  $A$ . If lower courts were judging in isolation, it would fall to the Supreme Court to review one case and set a universal doctrine of  $M$ . If lower courts are able to observe each other's decisions, though, they are able to take the workload off the Supreme Court by making the decision of  $M$  themselves.

**Proposition 5 (Learning Within Lower Courts)** *When homogeneous, unbiased lower courts can observe one another's decisions, the following is true in the unique equilibrium.<sup>8</sup>*

- $LC_I$  makes choices as he would before.
- $LC_{II}$  makes the following decisions.
  - He decides  $A$  if  $LC_I$  has decided  $A$  and he sees evidence  $(0, 0)$  or  $(-1, 0)$ , or if  $LC_I$  has decided  $M$  and he sees evidence  $(-1, 0)$  and  $(q^2 + 1 - q)\alpha < 1/2$ .
  - He decides  $B$  if  $LC_I$  has decided  $B$  and he sees evidence  $(0, 0)$  or  $(0, 1)$  or if  $LC_I$  has decided  $M$  and he sees evidence  $(0, 1)$  and  $(q^2 + 1 - q)\alpha < 1/2$ .
  - He decides  $M$  if he sees evidence  $-1, 1$ ; if  $LC_I$  has decided  $A$  and he sees evidence  $(0, 1)$  or if  $LC_I$  has decided  $B$  and he sees evidence  $(-1, 0)$ ; or if  $LC_I$  has decided  $M$  and he sees evidence  $(0, 0)$  or—if  $(q^2 + 1 - q)\alpha > 1/2$ —if he sees  $(-1, 0)$  or  $(0, 1)$ .
- The Supreme Court's behavior is as follows.
  - If both lower courts have decided  $A$ , or if both lower courts have decided  $B$ , or if both lower courts have decided  $M$ , the Supreme Court does not review.
  - If the sequence of decisions was  $M, A$  or  $M, B$ , the Supreme Court reviews the decision of  $M$  if  $c < L[1 - 2\frac{\alpha(1-q)^2}{\alpha(1-q)^2 + \alpha q^2 + (1-\alpha)}]$ .

---

<sup>8</sup>For this proposition, the assumption that the Supreme Court's loss from incorrect decisions is equal in magnitude (not only in balance) to lower courts' loss from incorrect decisions becomes critical. This is because a lower court now has an opportunity to change its decision on the basis of a prior lower court's ruling, so that differences in risk aversion between the Supreme Court and the Courts of Appeals can have important implications.

- *If the sequence of decisions was A, M or B, M, the Supreme Court reviews either if  $c < L$  and issues a binding doctrine of M.*

Therefore, the Supreme Court reviews much less often. It reviews *only* decisions of *M* in equilibrium, which are now much less ambiguous. When the decision of *M* is preceded by a decision of *A* or *B*, review is intended to increase uniformity (there is nothing to learn). When a decision of *M* is succeeded by a decision of *A* or *B*, the Supreme Court would only find it useful to review the decision of *M*.

In practice, lower courts do have opportunities to learn from one another (Baker and Malani 2015). When they see a case of first impression in their own circuit, they can and often do draw on reasoning from other circuits. But in many areas of the law, circuits' decisions are independent of one another, since common practice is to give consideration to other courts' decisions only to the extent a judge wishes. This presents a normative quandary: should judges make decisions independently, or should they learn from one another? In evaluating the Supreme Court's bankruptcy doctrine—widely agreed to be messy, bad law—Schwartz (2007) writes that the Courts of Appeals should behave as the Supreme Court does, and intelligently review one another's decisions:

If a unitary and competent high court had ultimate charge of the bankruptcy law, then each Federal circuit should develop those interpretations of the law that seem best to it. The highest court would then have the advantage of observing the likely full set of thoughtful positions when specifying what the bankruptcy law says. The Supreme Court's limited jurisdiction implies that the Court can have only a partial charge of the bankruptcy law, and ... the Court is not especially competent in the bankruptcy area. These considerations suggest that the circuits should view themselves as if they were state supreme courts when interpreting the Code. When the state courts function in fields with a national scope, such as contracts, they are influenced both by what they

think is best and by what other courts have done.

Normative suggestions on bankruptcy and contracts disputes encourage harmonization through intercircuit learning. Since there is no Supreme Court to resolve conflicts, lower courts should avoid creating them. However, the model presented here illustrates that sometimes courts *should* create conflicts—for example, if they receive messages  $(-1, 1)$  after the first circuit has not. Intercircuit splits are indications of legal development and are how circuits contribute knowledge. This result is consistent with Baker and Malani (2015), who consider lower courts who sequentially receive independent signals and can also observe the signals of prior lower courts.

## 5 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, Baker and Malani 2015). 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.

The theory's underlying premise—that lower courts' decisions are useful for doctrinal development, but only in concert—is consistent with the U.S. Supreme Court's own conception of its role. In the 1970s and 1980s, justices and policymakers became concerned about the Supreme Court's workload. The number of filings had ballooned while the number of signed opinions had remained relatively constant, resulting in an increasingly small percentage of cases being decided on the merits. In the ensuing discussions about the proper role for the Supreme Court, a debate arose: should the Supreme Court be in the business of articulating doctrine that would be self-evident to anyone who read the balance of evidence in the lower courts? A number of parties—including Chief Justice Warren Burger, the Hruska commission, and the Freund committee—called for the creation of various forms of an intercircuit panel that would resolve such straightforward questions.<sup>9</sup> In the terms of this model, an intercircuit panel would be tasked with certifying a doctrine of  $M$  after the lower courts had ruled  $(A, B)$ . This would leave the Supreme Court free to spend its time on the more difficult issues—deciding whether a set of decisions  $(A, M)$  should imply a doctrine of  $M$  or one of  $A$ . In other words, the notion of the Supreme Court being, by definition, an institution that learns about difficult issues, is pervasive among institutional designers and justices.

In *U.S. v. Mendoza* (464 U.S. 154, 1984), the Supreme Court unanimously held that the federal

---

<sup>9</sup>See Alsup and Salisbury (1984) and citations within for the history of this debate.

government cannot be stopped from relitigating a question of law, even if it has lost on that same issue in a previous lawsuit against a different party. Explaining the decision, Chief Justice Rehnquist wrote that “A rule allowing nonmutual collateral estoppel against the Government ... would substantially thwart the development of important questions of law by freezing the first final decision rendered on a particular legal issue. Allowing only one final adjudication would deprive this Court of the benefit it receives from permitting several courts of appeals to explore a difficult question before this Court grants certiorari.” If it were possible to collaterally estop the federal government, then failing to find evidence in favor of the government’s position in the first case would prevent any future attempts to find this information—which would severely hamper the Supreme Court’s opportunities to learn.

By illustrating how the Supreme Court would behave if it were trying to learn about doctrine, the model 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); Kastlelec (2007); Beim, Hirsch and Kastlelec (2014); 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.<sup>10</sup> Sim-

---

<sup>10</sup> 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.

ilarly, 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, 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

Alsup, William and Tracy L. Salisbury. 1984. "A Comment on Chief Justice Burger's Proposal for a Temporary Panel to Resolve Intercircuit Conflicts." *Hastings Constitutional Law Quarterly* 11:359–370.

- Austen-Smith, David and Jeffrey S. Banks. 1996. "Information Aggregation, Rationality, and the Condorcet Jury Theorem." *American Political Science Review* 90(1):34–45.
- Baker, Scott and Anup Malani. 2015. "Do Judges Actually Care About the Law? Evidence from Circuit Split Data." University of Chicago Working Paper.
- 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. 2014. "Whistleblowing and Compliance in the Judicial Hierarchy." *American Journal of Political Science* 58(4).
- 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 and Ryan Owens. 2012. "Looking Back to Move Forward: Quantifying Policy Predictions in Political Decision Making." *American Journal of Political Science* 56(4).
- 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.
- Callander, Steven and Bard Harstad. 2015. "Experimentation in Federal Systems." *Quarterly Journal of Economics* 130(2):951–1002.
- 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 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.
- Carrubba, Cliff, Barry Friedman, Andrew Martin and Georg Vanberg. 2012. "Who Controls the Content of Supreme Court Opinions?" *American Journal of Political Science* 56(2).
- Carrubba, Clifford J. and Tom S. Clark. 2012. "Rule Creation in a Political Hierarchy." *American Political Science Review* 106(3):622–643.
- 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.
- Daughety, Andrew F. and Jennifer F. Reinganum. 1999. "Stampede to judgement: pervasive influence and herding behavior by courts." *American Law and Economics Review* 1(1):158–189.
- 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, 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](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=789544).
- Hansford, Thomas G., James F. Spriggs and Anthony Stenger. 2010. "The Information Dynamics of Vertical Stare Decisis." Available at <http://ssrn.com/abstract=1580528>.

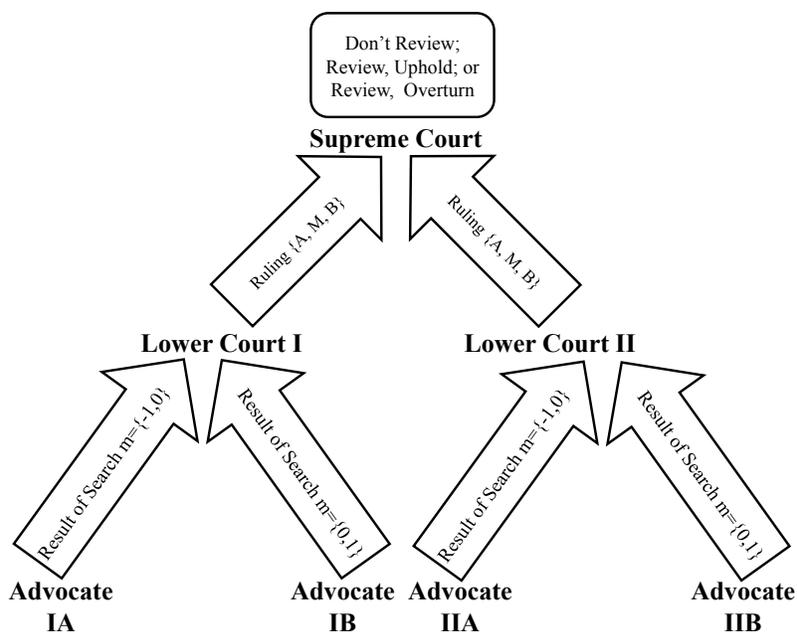
- Hettinger, Virginia A., Stefanie A. Lindquist and Wendy L. Martinek. 2004. "Comparing Attitudinal and Strategic Accounts of Dissenting Behavior on the U.S. Courts of Appeals." *American Journal of Political Science* 48:123–37.
- Kastellec, Jonathan P. 2007. "Panel Composition and Judicial Compliance on the U.S. Courts of Appeals." *Journal of Law, Economics & Organization* 23(2):421–441.
- 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).
- 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.
- 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.
- Schwartz, Alan. 2007. A Valuation of Collateral. In *Bankruptcy Law Stories*, ed. Robert Rasmussen. Foundation Press pp. 103–116.
- 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.
- Spitzer, Matthew and Eric Talley. 2012. "Left, Right, and Center: Strategic Information Acquisition and Diversity in Judicial Panels." *Journal of Law Economics & Organization* Forthcoming.

Summers, John S. and Michael J. Newman. 2011. "Towards a Better Measure and Understanding of U.S. Supreme Court Review of Courts of Appeals Decisions." *The United States Law Week* 80(393).

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.

Wasby, Stephen L. 2005. "How the Ninth Circuit Fares in the Supreme Court: The Intercircuit Conflict Cases." *Seton Hall Circuit Review* 1(1).



		$\theta$					
		Unbiased			Biased		
<i>Doctrine</i>	<b>A</b>	-1	0	1	-1	0	1
	<b>M</b>	0	-L	-1	0	-L	-1
	<b>B</b>	-L	0	-L	-L	0	-L
		-1	-L	0	-1	-1	0

Figure 1: **Upper panel:** Play of game. **Lower panel:** Judges' preferences over doctrine, conditional on the state of the world  $\theta$ . 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 panel:* For judges biased against  $B$ , wrongly choosing  $B$  is more costly than wrongly choosing  $A$ .